Constraint and Contingency Pervade the Emergence of Novel Phenotypes in Complex Metabolic Systems

Hosseini, Sayed-Rzgar ; Wagner, Andreas

Abstract: An evolutionary constraint is a bias or limitation in phenotypic variation that a biological system produces. We know examples of such constraints, but we have no systematic understanding about their extent and causes for any one biological system. We here study metabolisms, genomically encoded complex networks of enzyme-catalyzed biochemical reactions, and the constraints they experience in bringing forth novel phenotypes that allow survival on novel carbon sources. Our computational approach does not limit us to analyzing constrained variation in any one organism, but allows us to quantify constraints experienced by any metabolism. Specifically, we study metabolisms that are viable on one of 50 different carbon sources, and quantify how readily alterations of their chemical reactions create the ability to survive on a novel carbon source. We find that some metabolic phenotypes are much less likely to originate than others. For example, metabolisms viable on D-glucose are 1835 times more likely to give rise to metabolisms viable on D-fructose than on acetate. Likewise, we observe that some novel metabolic phenotypes are more contingent on parental phenotypes than others. Biochemical similarities among carbon sources can help explain the causes of these constraints. In addition, we study metabolisms that can be produced by recombination among 55 metabolisms of different bacterial strains or species, and show that their novel phenotypes are also contingent on and constrained by parental genotypes. To our knowledge, our analysis is the first to systematically quantify the incidence of constrained evolution in a broad class of biological system that is central to life and its evolution.

DOI: https://doi.org/10.1016/j.bpj.2017.06.034

Biophysical Journal, Volume *113*

Supplemental Information

Constraint and Contingency Pervade the Emergence of Novel Phenotypes in Complex Metabolic Systems

Sayed-Rzgar Hosseini and Andreas Wagner

1. Supplementary Methods:

S1: Genome-scale metabolic networks and their phenotypic representations

Similar to our previous work describing the procedures used here (1), and following common practice in metabolic systems biology (2–4), we represent an organism's metabolic genotype as the set of genomically encoded (enzyme-catalyzed) biochemical reactions proceeding inside the organism. This metabolic genotype specifies a metabolism or metabolic network, a network of chemical reactions encoded by the genotype. A metabolic reaction network enables an organism to extract energy and produce small biomass building blocks, such as amino acids, from extracellular nutrients. Inference of this genotype from genomic and biochemical information has been successful for multiple organisms (5, 6).

Any one metabolic reaction network contains a subset of the "reaction universe" of all biochemical reactions that take place in prokaryotes (See text S2). We have curated a representation of this universe, which comprises 5,906 reactions and is based on current metabolic knowledge (7–10). We represent an organism's metabolic genotype as a binary vector of length 5,906. Each entry of this vector corresponds to a given reaction in the reaction universe, and is equal to one if the corresponding reaction is present in the metabolic network, and zero otherwise. Thus, each genotype can be thought of as a single member of a vast space of all possible metabolic networks, which contains 2^{5906} distinct genotypes.

We define the phenotype of a given metabolic genotype based on its viability in 50 distinct minimal environments that differ only in the carbon source they harbor (See Text S3). We consider that a genotype is *viable* on a given carbon source, if it can produce all essential biomass precursor molecules from the given carbon source, and we use Flux Balance Analysis (FBA, See text S4) to determine viability (11). We represent the phenotype of a given metabolic genotype as a binary vector of length 50. Each entry of this vector corresponds to a given carbon source, and it is equal to one if the genotype is viable on this carbon source, and zero otherwise.

S2: Reaction universe

The reaction universe we curated is a set of metabolic reactions in which each reaction is known to occur in some prokaryotic organisms. For the curation of this universe, we used data from the LIGAND database (7, 8) of the Kyoto Encyclopedia of Genes and Genomes (9). Briefly, the LIGAND database, which is comprised of the REACTION and the COMPOUND databases, provides information on reactions, associated stoichiometric information, chemical compounds involved in a reaction, and the Enzyme Classification (E.C.) identifier of each

reaction. From the REACTION and the COMPOUND databases we excluded (i) all reactions involving polymer metabolites of unspecified numbers of monomers, or general polymerization reactions with uncertain stoichiometry, (ii) reactions involving glycans, due to their complex structure, (iii) reactions with unbalanced stoichiometry, and, (iv) reactions involving complex metabolites without chemical information about their structure (10). Moreover, we do not consider unknown reactions, and we also do not take into account spontaneous reactions, or reactions that depend on external stimuli. The published *E. coli* metabolic model (iAF1260) consists of 1397 non-transport reactions (12). We merged all reactions in the *E. coli* model with the reactions in the KEGG dataset, and retained only the unique (non-duplicate) reactions. This resulted in a universe of reactions consisting of 682 transport, 5,906 non-transport reactions and 5030 metabolites. The reaction universe is available online (https://github.com/rzgar/EMETNET/tree/master/UNIVERSE).

S3: Chemical environments

We consider 50 minimal growth environments, each of which includes oxygen, ammonium, inorganic phosphate, sulfate, sodium, potassium, cobalt, iron (Fe^{2+} and Fe^{3+}), protons, water, molybdate, copper, calcium, chloride, magnesium, manganese, zinc, and a specific carbon source. Importantly, to represent different chemical environments, we vary the carbon source while keeping all other nutrients constant. We consider a metabolic network viable on a given carbon source, if it can synthesize all essential biochemical precursors when this carbon source is provided as the sole carbon source in the minimal medium just described.

We used 50 carbon sources for our analysis of randomly sampled metabolic networks, including the following 27 glycolytic carbon sources: D-glucose, D-glucose 6-phosphate, trehalose, maltose, lactose, D-fructose 6-phosphate, D-fructose, D-mannose, D-mannitol, D-glucose 1-phosphate, D-sorbitol, maltotriose, D-allose, D-ribose, D-xylose, D-gluconate, 5-dehydro-D-gluconate, L-rhamnose, L-fucose, L-arabinose, L-lyxose, D-galactose, melibiose, D-galactonate, N-acetyl-D-glucosamine, N-acetyl-D-mannosamine, N-acetylneuraminate.

In addition, we used the following 20 gluconeogenic carbon sources: pyruvate, L-alanine, L-lactate, D-alanine, D-malate, acetate, L-serine, L-malate, D-serine, glycine, glycolate, L-aspartate, succinate, fumarate, 2-oxoglutarate, D-galacturonate, D-galactarate, D-glucarate, L-galactonate, D-glucoronate. And we used the following three nucleosides as carbon sources: adenosine, deoxyadenosine, inosine.

To study the emergence of novel phenotypes in 55 prokaryotic metabolic networks from the BiGG database (13) (see methods section 2.4 in the main text), we used the following 30 carbon sources on which none of the 55 metabolic networks are predicted to be viable: Biotin, riboflavin, folate, pimelate, urea, carbonic acid, bicarbonate, methanol, trimethylamine, D-

methionine, glycine betaine, gamma-butyrobetaine, choline, L-phenylalanine, L-leucine, L-tyrosine, L-methionine, thiamin, 6-diaminoheptanedioate, (R)-pantothenate, spermidine, taurine, isocytosine, protoheme, nicotinamide adenine dinucleotide, L-fucose 1-phosphate, dimethyl-sulfide, L-carnitine, dimethyl sulfoxide, and 1,5-diaminopentane.

S4: Flux balance analysis

Flux balance analysis (FBA) is a computational method that is widely used for the quantitative analysis and modeling of metabolic networks (11). Based on the stoichiometric coefficients of the metabolites participating in the reactions of a given metabolic network, FBA predicts the metabolic flux through each reaction. Stoichiometric coefficients are stored in a stoichiometric matrix S, which is of dimension $m \times n$, where m and n, denote the number of metabolites and the number of reactions in a metabolic network. FBA constrains the flux through each reaction based on the assumption that a metabolic network is in a steady state where metabolite concentrations do not change, i.e., $Sv = 0$, where v is the vector of metabolic fluxes v_i through reaction i. The solutions of the equation $Sv = 0$, that is, the null space of matrix S, comprises all flux vectors that are allowable in steady state. The null space is further constrained by physicochemical information regarding the maximum and minimum possible fluxes through each reaction. FBA relies on an optimization procedure called linear programming to identify those among the allowable flux vector(s) that maximize an objective function Z. This task can be formulated as finding a flux vector v^* with the property

$$v^* = \max_v Z(v) = \max_v \{ c^T v \mid Sv = 0,\ a \leq v \leq b \},$$

where the vector c contains a set of scalar coefficients representing the maximization criterion, and each entry a_i and b_i of vectors a and b, indicates the minimally and maximally possible flux through reaction i. The vector c represents the proportions of each small biomass molecule in a cell's biomass. Therefore v^* maximizes the biomass growth flux, that is, the rate at which a metabolic network can produce biomass (11). Here we use FBA to predict qualitatively whether a given metabolic network is viable in a given environment, and we consider a metabolic network viable if it can produce all essential biomass precursors. More precisely, FBA predicts a metabolic network as viable on a given environment, if its biomass flux rate exceeds 0.001 $1/h$. In a free-living bacterium like *E.coli,* there are approximately 60 such molecules including 20 amino acids, DNA, and RNA precursors, lipids and cofactors. We used the biomass composition of the *E. coli* metabolic model iAF1260 to define the vector c (12). Moreover, we used the packages CPLEX (11.0, ILOG; http://www.ilog.com/) and CLP (1.4, Coin-OR; https://projects/coin-or.org/Clp) to solve the linear programming problem of FBA.

The major limitation of FBA is that it neglects regulatory constraints that can arise through suboptimal expression or regulation of enzymes. Newly horizontally transferred genes cannot easily establish regulatory interactions with their host genes, and it may thus take considerable adaptive evolution until they become expressed at a maximal or optimal level (14). Such regulatory constraints would be especially important if we focused on quantitative predictions of biomass growth (15). However, we use FBA solely for qualitative prediction of viability. This focus on qualitative phenotypes is biologically sensible. The reason is that many organisms grow slowly in their native environment (16, 17), implying that regulation for maximal biomass production is far from universal. Moreover, we note that regulatory constraints can easily be broken in evolution, even on the short time scales of laboratory evolution experiments (15, 18, 19).

S5: Generation of random metabolic networks

We here employ a previously described *in silico* process which relies on Markov Chain Monte Carlo (MCMC) random walks to generate metabolic networks that comprise random sets of metabolic reactions that are viable on a given carbon source (10, 20). This procedure can produce metabolic networks that are sampled uniformly from the set of all metabolic networks viable on a given carbon source (10, 20). Briefly, in each step of such a random walk we perform a reaction swap, defined as altering a metabolic network by adding a randomly chosen reaction from the reaction universe, and then deleting a reaction randomly chosen from the set of reactions present in the metabolic network. If the reaction swap disrupts the metabolic network's viability on the given carbon source (as determined by FBA) we reject it, and perform another reaction swap until we find a swap that does not disrupt viability. This procedure also ensures that the total number of reactions remains constant. For the MCMC method to produce random samples of metabolic networks, it is essential to carry out enough reaction swaps to "erase" the random walker's similarity to the initial metabolic network. Previously, it has been shown that 3×10^3 reaction swaps are sufficient for this purpose (10, 20). Each of our random walks starts from *E. coli*'s metabolic network and performs 10^4 reaction swaps before storing the final metabolic network for further analysis. We used 10^4 independent random walks conducted in this way to create 10^4 random metabolic networks viable on each of the 50 carbon sources.

S6: Generation of parental metabolic network pairs

Some of our analyses required us to recombine pairs of "parental" metabolic networks with particular features, such as being viable on a specific carbon source (and only on that carbon source), or having a given genotypic distance (D), defined as the number of reactions differing

between the parents. Generating parents with a given genotypic distance (D) is not straightforward, because the random metabolic networks generated by MCMC sampling generally have genotypic distances sufficiently large ($D \approx 2{,}000$) to be biologically unrealistic for modeling frequently recombining prokaryotic genomes. To create less distant metabolic network pairs, we took an MCMC random walk approach. It revolves around a reaction-swapping random walk starting with a pair of randomly chosen metabolic networks from our sample of 10^4 sampled metabolic networks that are exclusively viable on a given carbon source. In each step of this random walk, we subjected each parental metabolic network to a reaction swap, and we accepted each reaction swap if it (*i*) preserved the original phenotype, and (*ii*) did not increase the genotypic distance of the two metabolic networks after the swap, otherwise we rejected the reaction swap. We continued this procedure until the genotypic distance between the metabolic networks became equal to a desired distance D. We note that this procedure is very time-consuming when applied to the thousands of parents we study here.

Finally, to generate parental metabolic networks with a given number of reactions, we started from a random viable metabolic network generated by MCMC sampling, as described in the text S5. All such metabolic networks have the same number of reactions as *E.coli* (2,079). We then applied a sequence of individual and random reaction deletions, where we required that each deletion preserve viability, until the network had reached the desired size.

S7. Estimation of the metabolic distance between carbon sources

For each pair of carbon sources (C_i, C_j), we calculated metabolic distance with two different approaches, a direct approach that is based on the shortest path between carbon sources in substrate graph (21), and an indirect approach that is based on carbon source-dependent superessentiality of metabolic reactions in metabolic networks (22).

The first approach relies on the substrate graph of a metabolic network, in which vertices correspond to metabolites. Two metabolites are linked via an edge, if the metabolites participate in the same metabolic reactions as either a substrate or a product. From this substrate graph we excluded currency metabolites, which are metabolites that transfer small chemical groups, and are involved in many reactions (23). Specifically, we excluded protons, H_2O, ATP (adenosine triphosphate), ADP (adenosine diphosphate), AMP (adenosine monophosphate), NADP(H) (nicotinamide adenosine dinucleotide diphosphate), NAD(H) (nicotinamide adenosine dinucleotide), and P_i (inorganic phosphate), CoA (coenzyme A), hydrogen peroxide, ammonia, ammonium, bicarbonate, GTP (guanosine triphosphate), GDP (guanosine diphosphate), and PP_i (inorganic diphosphate) that occurred in both the cytoplasmic and periplasmic compartments. In addition, we excluded oxidized and reduced

forms of cofactors such as quinone, ubiquinone, glutathione, thioredoxin, flavodoxin and flavin mononucleotide. For all metabolic networks viable on C_i, we measured the shortest path in the substrate graph between C_i and any other $C_j, j \neq i$ using Dijkstra's algorithm (24). Then, we considered the average shortest path between C_i and C_j among metabolic networks viable on C_i as the metabolic distance between C_i and C_j.

In the second approach, we take advantage of the fact that metabolic reactions show varying degrees of essentiality among different metabolic networks that are viable on the same carbon sources. Any one reaction can be essential in one such network and inessential in another, depending on which reactions and pathways are present in the network. One can quantify a reaction's degree of essentiality in randomly sampled viable networks via a "superessentiality index", defined as the fraction of metabolic networks in which the reaction is essential for viability on a given carbon source (22). Highly superessential reactions are essential in most random viable networks, and cannot be by-passed easily by alternative metabolic pathways. We first computed the superessentiality index of each reaction on each carbon source C_i, and assembled this information into a superessentiality vector. Each element of this vector corresponds to one of the 5,906 reactions in the reaction universe, and contains the fraction of random viable metabolic networks in which the reaction is essential for viability on C_i. We then computed the Euclidian distance between the superessentiality vectors for all pairs of carbon sources C_i and C_j as a proxy for metabolic distance between the two carbon sources.

S8: Distance measure between carbon sources based on superessential reactions

In the second approach, we take advantage of the fact that metabolic reactions show varying degrees of essentiality among different metabolic networks that are viable on the same carbon sources. Any one reaction can be essential in one such network and inessential in another, depending on which reactions and pathways are present in the network. One can quantify a reaction's degree of essentiality in randomly sampled viable networks via a "superessentiality index", defined as the fraction of metabolic networks in which the reaction is essential for viability on a given carbon source (22). Highly superessential reactions are essential in most random viable networks, and cannot be by-passed easily by alternative metabolic pathways. We first computed the superessentiality index of each reaction on each carbon source C_i, and assembled this information into a superessentiality vector. Each element of this vector corresponds to one of the 5,906 reactions in the reaction universe, and contains the fraction of random viable metabolic networks in which the reaction is essential for viability on C_i. We then computed the Euclidian distance between the superessentiality vectors for all pairs of carbon sources C_i and C_j as a proxy for metabolic distance between the two carbon sources.

Previous work showed that highly superessential reactions are more likely to be involved in metabolic innovation (1). We thus also wanted to compute a biochemical distance measure of carbon sources based on this index. To this end, we computed, for each carbon source, the superessentiality index of all reactions belonging to the reaction universe, which yields a superessentiality vector of length 5,906. We then computed the Euclidian distance between the superessentiality vectors for all pairs of carbon sources C_i and C_j as a proxy for the biochemical distance between the two carbon sources. Fig. S27a shows that the number of innovative offspring, which are generated by recombination between parents viable on C_i, and gain viability on a given carbon source C_j is significantly correlated with the Euclidian distance between the superessentiality vectors for (C_i, C_j) (Pearson r = -0.3935, and $P < 10^{-83}$).

S9: Random metabolic networks and erroneous energy generating cycles

A recent study by Fritzemeier et al. showed that most of the published genome-scale metabolic networks include thermodynamically impossible energy-generating cycles (EGCs), which are capable of charging energy metabolites without nutrient consumption (25). It showed that these EGCs can artificially inflate biomass flux by 25% and could be particularly problematic in evolutionary simulations, which involves incorporation of foreign metabolic reactions from other species.

We applied the approach of Fritzemeier et al., to identify EGCs in metabolic networks (25), using 15 different energy dissipation reactions (EDRs) for each of the 15 different types of energy metabolites in the cell. (See https://doi.org/10.1371/journal.pcbi.1005494.s002 for complete information on these reactions). We maximized one energy dissipation reaction flux v_d at a time, while preventing all influx of external nutrients into the model. The problem can be mathematically expressed as follows:

$$\max v_d \text{ subject to:}$$
$$Sv = 0$$
$$\forall i \notin E: v_i^{min} \leq v_i \leq v_i^{max}$$
$$\forall i \in E: v_i = 0$$

where S is the stoichiometric matrix describing a metabolic system, v is the vector of all metabolic fluxes, d is the index of one of the energy dissipation reactions, v^{min} and v^{max} are vectors of lower and upper reaction bounds, and E is the set of indices of all exchange reactions. An optimal value v_d^* for this optimization with $v_d^* > 0$ for at least one of the energy dissipation reactions demonstrates the existence of at least one EGC in the corresponding metabolic network.

Using this approach, we first determined that the initial *E. coli* metabolic network with 2079 reactions (12) from which we started most of our MCMC sampling had no EGCs. However, we found that 97.3% and 97.8% of our randomly sampled metabolic networks viable on

glucose and acetate, respectively, harbored at least one EGC.

To determine whether these EGCs artificially inflated the number of innovative offspring, we sampled EGC-free parental metabolic networks. To do so, we modified our MCMC approach such that each sampled metabolic network not only retained viability in a given environment, but was also EGC-free. To fulfill these goals, we required that each step (reaction swap) in our MCMC sampling preserved viability on a given carbon source, and did not introduce an EGC (checked by the EGCs identification approach described above). Using this approach, we generated 1,000 pairs of EGC-free metabolic networks viable exclusively on glucose, and 1,000 pairs of EGC-free networks viable only on acetate. We then generated 1,000 recombinant offspring from each pair. Recombination between EGC-free metabolisms viable exclusively on glucose resulted in 29,941 innovative offspring, only 7.41% fewer than the corresponding number for EGC-containing metabolisms (32,338). Likewise, we observed 46,941 innovative offspring emerging from EGC-free parental metabolisms viable exclusively on acetate, 5.57% fewer than the corresponding number for EGC-containing metabolisms (49,708). Thus, removing EGCs slightly reduces the incidence of innovation (figure S30). Importantly, the patterns of relative constraints remain almost exactly unchanged (figure S31).

Fritzemeier et al. showed that EGCs could artificially increase the biomass rate of metabolic networks by 25% (25). However, figure S32 indicates that the majority of viable networks we study already have a biomass flux considerably larger than our threshold of viability, so reducing their biomass production rate by 25% will not result in a viability loss for most metabolisms, which is why excluding EGCs does not substantially reduce the emergence of novel phenotypes.

Supporting References:

1. Hosseini, S.-R., O. Martin, and A. Wagner. 2016. Phenotypic innovation through recombination in genome-scale metabolic networks. Proc. R. Soc. B. .
2. Edwards, J.S., R.U. Ibarra, and B.O. Palsson. 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat. Biotechnol. 19: 125–30.
3. Edwards, J.S., and B.O. Palsson. 1999. Systems properties of the Haemophilus influenzae Rd metabolic genotype. J. Biol. Chem. 274: 17410–6.
4. Lewis, N.E., H. Nagarajan, and B.O. Palsson. 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10: 291–305.
5. Feist, A.M., and B.Ø. Palsson. 2008. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat. Biotechnol. 26: 659–67.
6. McCloskey, D., B.Ø. Palsson, and A.M. Feist. 2013. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. Mol. Syst. Biol.

9: 661.

7. Goto, S., T. Nishioka, and M. Kanehisa. 2000. LIGAND: chemical database of enzyme reactions. Nucleic Acids Res. 28: 380–2.

8. Goto, S., Y. Okuno, M. Hattori, T. Nishioka, and M. Kanehisa. 2002. LIGAND: database of chemical compounds and reactions in biological pathways. Nucleic Acids Res. 30: 402–4.

9. Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa. 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 38: D355-60.

10. Matias Rodrigues, J.F., and A. Wagner. 2009. Evolutionary plasticity and innovations in complex metabolic reaction networks. PLoS Comput. Biol. 5: e1000613.

11. Orth, J.D., I. Thiele, and B.Ø. Palsson. 2010. What is flux balance analysis? Nat. Biotechnol. 28: 245–8.

12. Feist, A.M., C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.Ø. Palsson. 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3: 121.

13. King, Z.A., J. Lu, A. Dräger, P. Miller, S. Federowicz, J.A. Lerman, A. Ebrahim, B.O. Palsson, and N.E. Lewis. 2015. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. : gkv1049-.

14. Lercher, M.J., and C. Pál. 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol. Biol. Evol. 25: 559–67.

15. Ibarra, R.U., J.S. Edwards, and B.O. Palsson. 2002. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature. 420: 186–9.

16. Vieira-Silva, S., and E.P.C. Rocha. 2010. The systemic imprint of growth and its uses in ecological (meta)genomics. PLoS Genet. 6: e1000808.

17. Kirschner, D., and S. Marino. 2005. Mycobacterium tuberculosis as viewed through a computer. Trends Microbiol. 13: 206–11.

18. Fong, S.S., and B.Ø. Palsson. 2004. Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. Nat. Genet. 36: 1056–8.

19. Fong, S.S., J.Y. Marciniak, and B.O. Palsson. 2003. Description and Interpretation of Adaptive Evolution of Escherichia coli K-12 MG1655 by Using a Genome-Scale In Silico Metabolic Model. J. Bacteriol. 185: 6400–6408.

20. Samal, A., J.F. Matias Rodrigues, J. Jost, O.C. Martin, and A. Wagner. 2010. Genotype networks in metabolic reaction spaces. BMC Syst. Biol. 4: 30.

21. Wagner, A., and D.A. Fell. 2001. The small world inside large metabolic networks. Proc. R. Soc. B Biol. Sci. 268: 1803–1810.

22. Barve, A., J.F.M. Rodrigues, and A. Wagner. 2012. Superessential reactions in metabolic networks. Proc. Natl. Acad. Sci. U. S. A. 109: E1121-30.

23. Ma, H.-W., and A.-P. Zeng. 2003. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics. 19: 1423–30.

24. Hopcroft, J., and R. Tarjan. 1973. Algorithm 447: efficient algorithms for graph manipulation. Commun. ACM. 16: 372–378.

25. Fritzemeier, C.J., D. Hartleb, B. Szappanos, B. Papp, M.J. Lercher, and G. Fekete. 2017. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. PLOS Comput. Biol. 13: e1005494.

2. Supplementary Figures

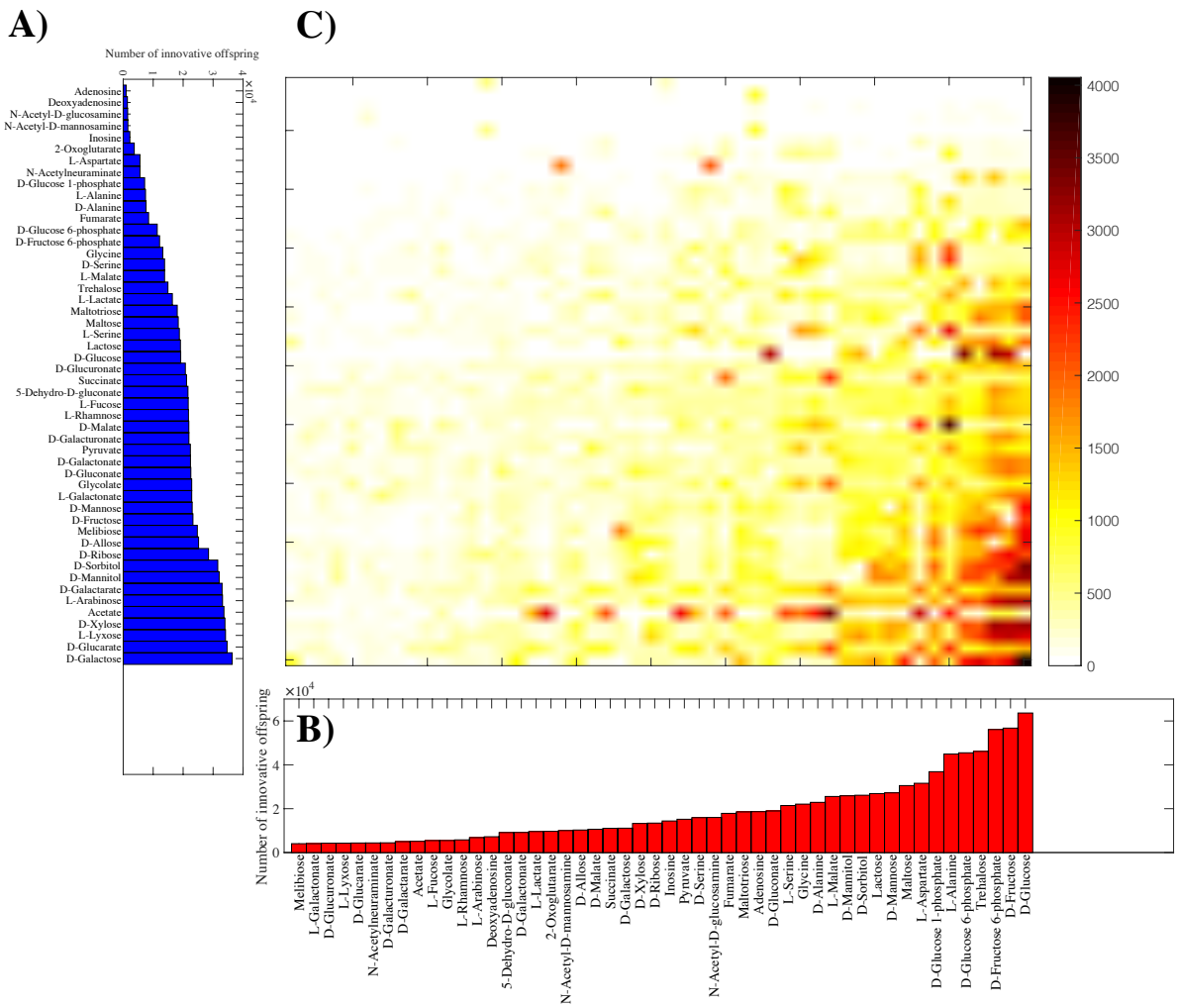


Figure S1: Recombination can create all 50 carbon-use phenotypes considered here (n = 20). A) The horizontal axis shows the number of innovative recombinant offspring (out of one million offspring) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis. This number varies by a factor 37, ranging from 977 on Adenosine to 356,378 on D-galactose. **B)** Number of innovative recombinants (per million offspring) gaining viability on the novel carbon source specified on the x-axis. This number varies by a factor 15, ranging from 4,042 on melibiose to 63,634 on D-glucose. **C)** Number of innovative recombinants (per million offspring, color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified in panel A, which have gained viability on the novel carbon source specified in panel B In these analyses, parental metabolic networks contain $\|G\|$ = 2,079 reactions, the same as in the *E.coli* metabolic network, and they differ in D = 100 reactions. Moreover, n = 20 reactions are swapped between parental metabolic networks in a recombination event.

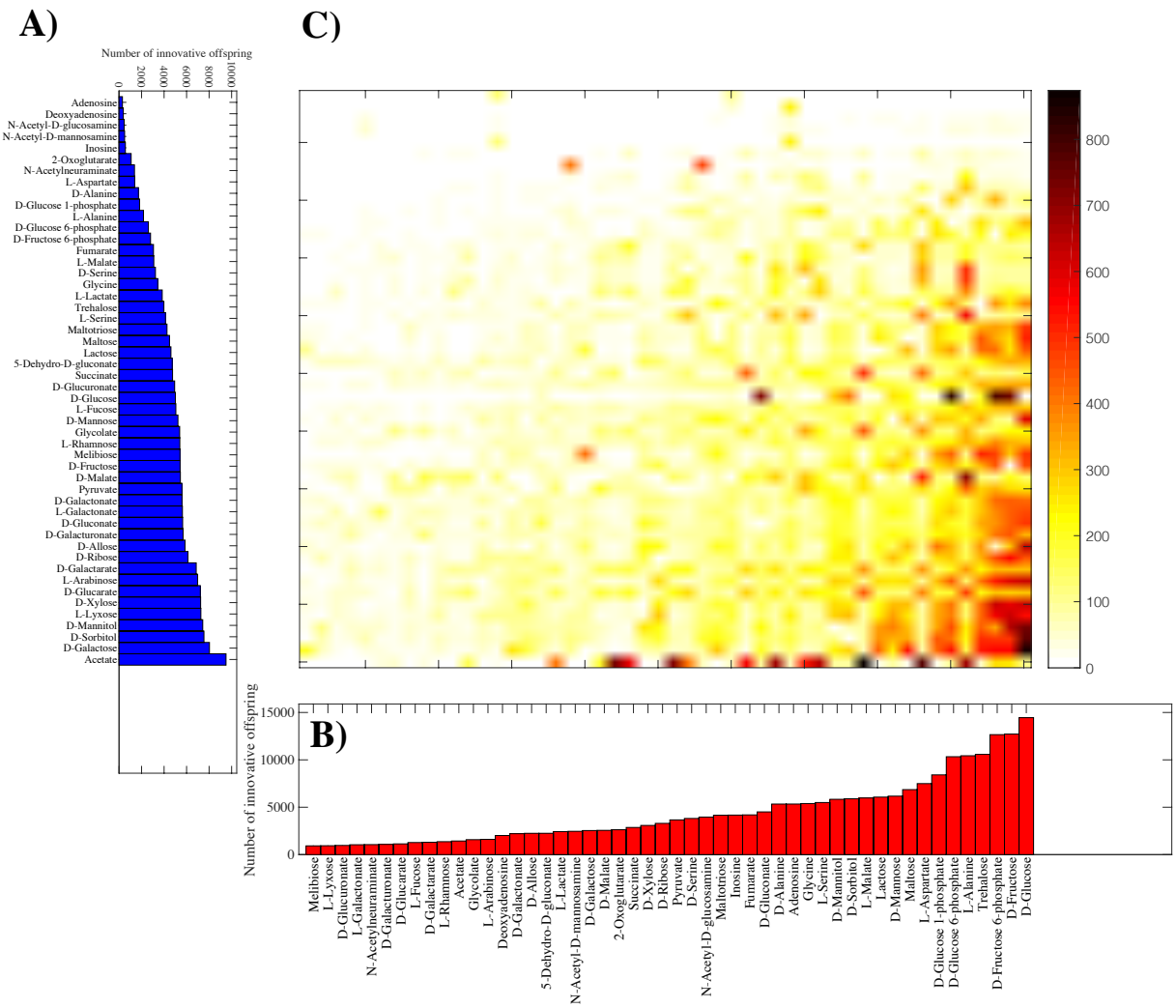


Figure S2: Recombination can create all 50 carbon-use phenotypes considered here (*n* = 30). A) The horizontal axis shows the number of innovative recombinant offspring (out of one million offspring) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis. This number varies by a factor 32, ranging from 299 on adenosine to 9,503 on acetate. **B)** Number of innovative recombinants (per million offspring) gaining viability on the novel carbon source specified on the *x*-axis. This number varies by a factor 16, ranging from 923 on melibiose to 14,452 on D-glucose. **C)** Number of innovative recombinants (per million offspring, color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified in panel A, which have gained viability on the novel carbon source specified in panel B. In these analyses, parental metabolic networks contain $\|G\| = 2{,}079$ reactions, the same as in the *E.coli* metabolic network, and they differ in $D = 100$ reactions. Moreover, $n = 30$ reactions are swapped between parental metabolic networks in a recombination event.

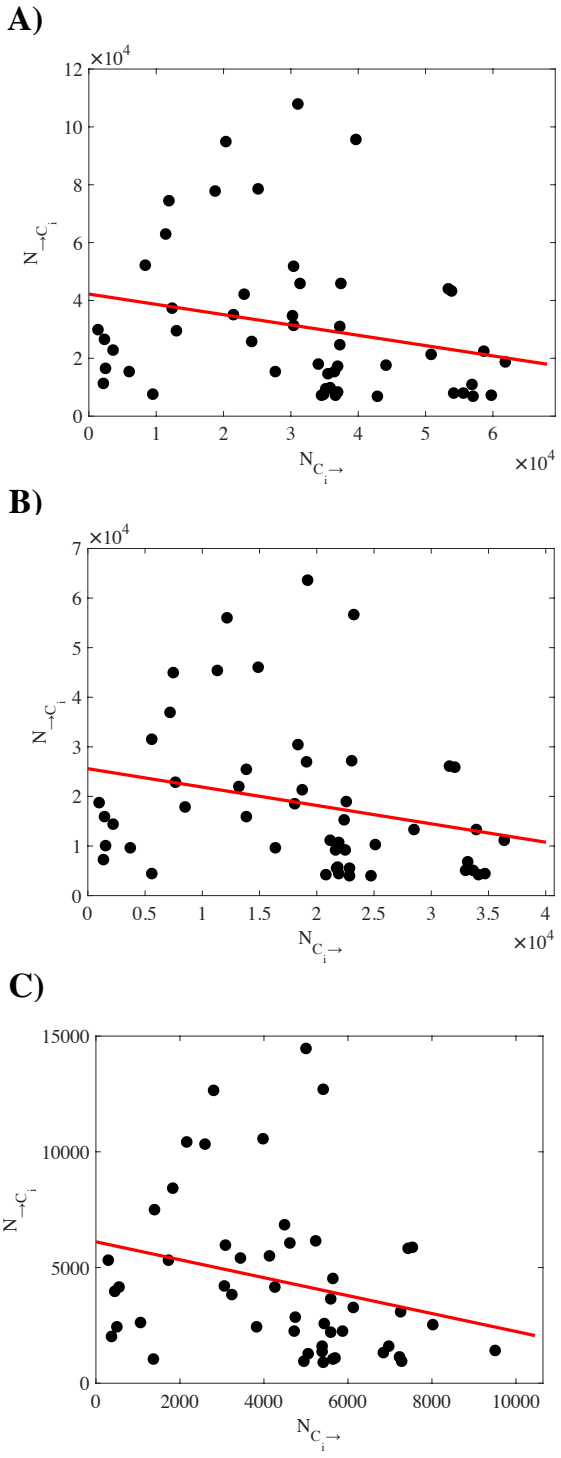
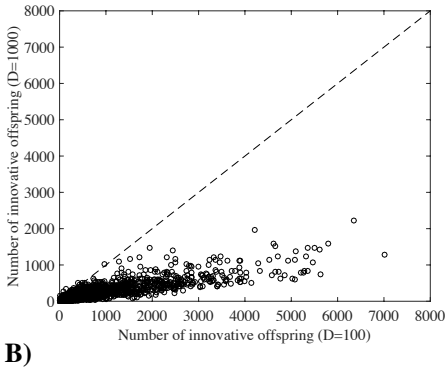
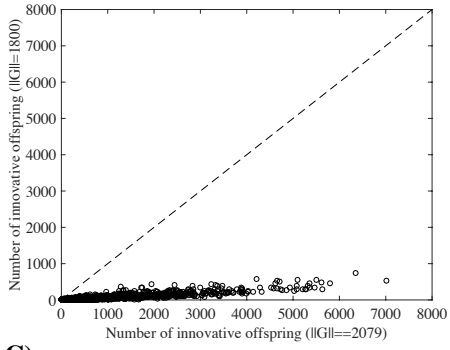


Figure S3: Negative correlation between $(N_{C_i\rightarrow})$ **and** $(N_{\rightarrow C_i})$**.** Each circle corresponds to a given carbon source C_i. The vertical axis shows $(N_{C_i\rightarrow})$, the number of metabolic innovations emerging from parents viable on carbon source C_i. The horizontal axis shows $(N_{\rightarrow C_i})$, the number of innovations leading to viability on C_i. There is a negative correlation between $(N_{C_i\rightarrow})$ and $(N_{\rightarrow C_i})$, regardless of the number (n) of reactions exchanged: **A)** (n = 10, Pearson r = -0.239, P < 0.093), **B)** (n = 20, Pearson r = -0.248, P < 0.082), **C)** (n = 30, Pearson r = -0.256, P < 0.073). For all analyses the genotypic distance between parents is D = 100.

A)



B)



C)

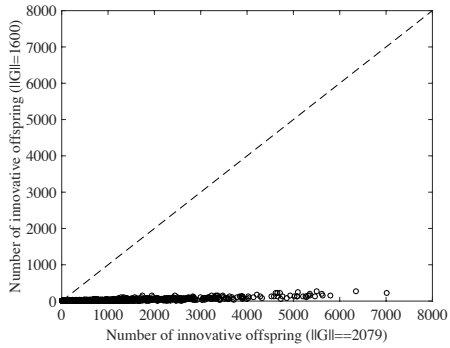


Figure S4: Fewer innovative offspring at higher genotypic distance (D) and smaller metabolic network size $\|G\|$. Each circle corresponds to a pair of carbon sources (C_i, C_j) and shows the number of innovative offspring gaining viability on C_j, which are generated by recombination between parents viable on carbon source C_i. The horizontal axis specifies the number of innovative offspring where parents have genetic distance D = 100, and metabolic network size $\|G\|$ = 2,079. The vertical axes provide the same information, but for parents with A) genotypic distance D = 1,000, and metabolic network size $\|G\|$ = 2,079 reactions, B) genotypic distance D = 100, and metabolic network size $\|G\|$ = 1,800 reactions, and C) genotypic distance D = 100, and metabolic network size $\|G\|$ = 1,600 reactions. The dashed diagonal lines correspond to the identity line ($y = x$). Note that in all three panels, most or all data lie below this line, indicating that higher parental genotypic distance and lower metabolic network size lead to fewer innovative offspring for almost all carbon source pair.

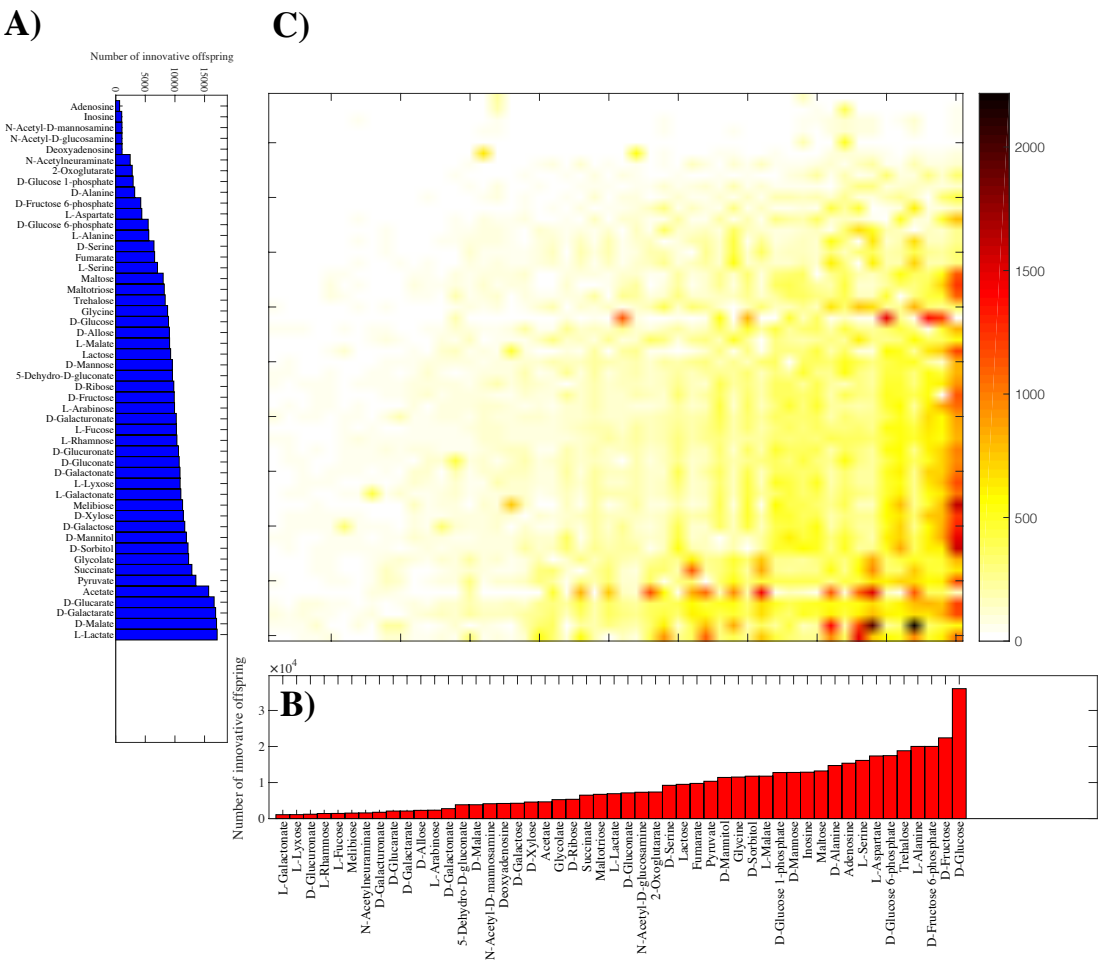


Figure S5: Recombination can create all 50 carbon-use phenotypes considered here (*D* = 1,000). A) The horizontal axis shows the number of innovative recombinant offspring (out of one million offspring) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis. This number varies by a factor 25, ranging from 662 on adenosine to 17,132 on L-lactate. **B)** Number of innovative recombinants (per million offspring) gaining viability on the novel carbon source specified on the *x*-axis. This number varies by a factor 33, ranging from 1081 on L-galactonate to 36,051 on D-glucose. **C)** Number of innovative recombinants (per million offspring, color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified in panel A, which have gained viability on the novel carbon source specified in panel B. In these analyses, parental metabolic networks contain $||G||$ = 2,079 reactions, the same number as the *E.coli* metabolic network, and they differ in D = 1,000 reactions. Moreover, n = 10 reactions are swapped between parental metabolic networks in a recombination event.

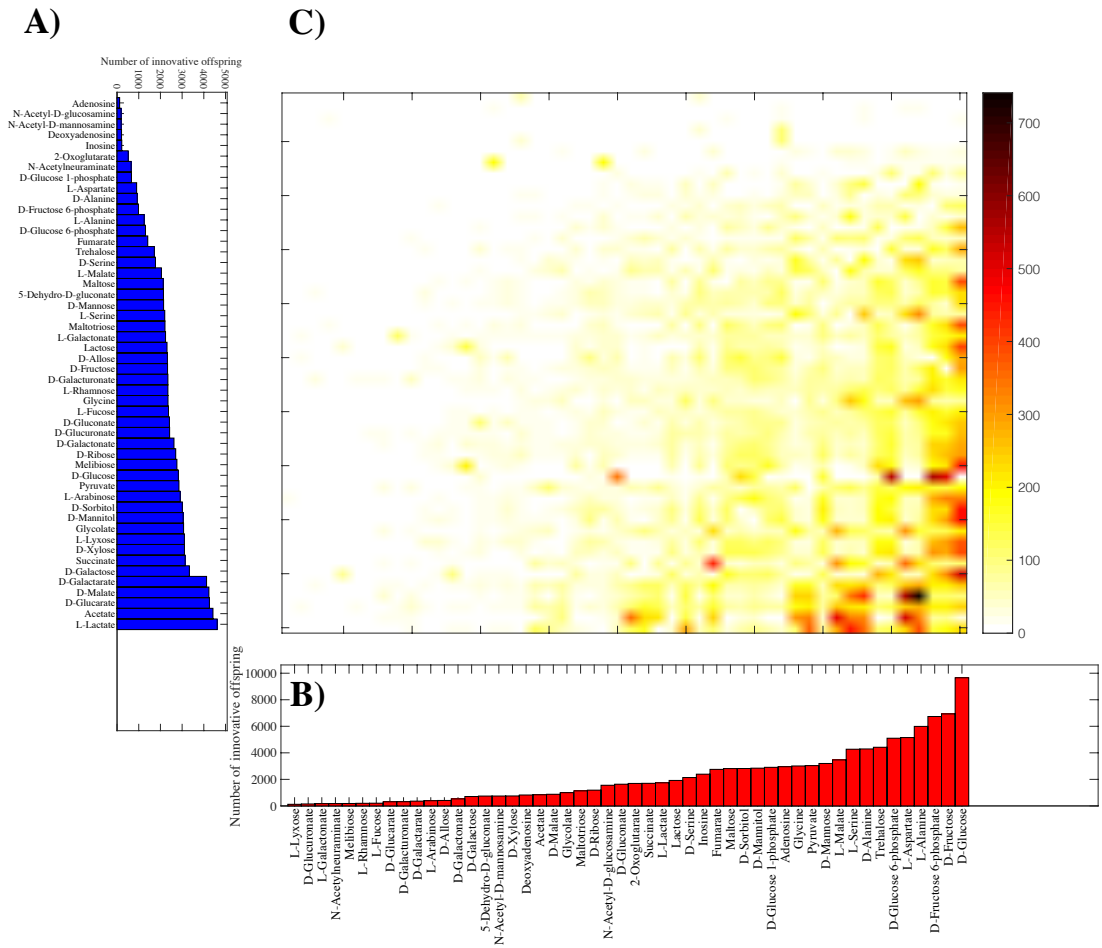


Figure S6: Recombination can create all 50 carbon-use phenotypes considered here ($\|G\|$ = 1800). A) The horizontal axis shows the number of innovative recombinant offspring (out of one million offspring) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis. This number varies by a factor 38, ranging from 120 on adenosine to 4,616 on L-lactate. **B)** Number of innovative recombinants (per million offspring) gaining viability on the novel carbon source specified on the *x*-axis. This number varies by a factor 79, ranging from 122 on L-lyxose to 9,657 on D-glucose. **C)** Number of innovative recombinants (per million offspring, color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified in panel A, which have gained viability on the novel carbon source specified in panel B. In these analyses, parental metabolic networks contain $\|G\|$ = 1,800 reactions and differ in D = 100 reactions. Moreover, n = 10 reactions are swapped between parental metabolic networks in a recombination event.

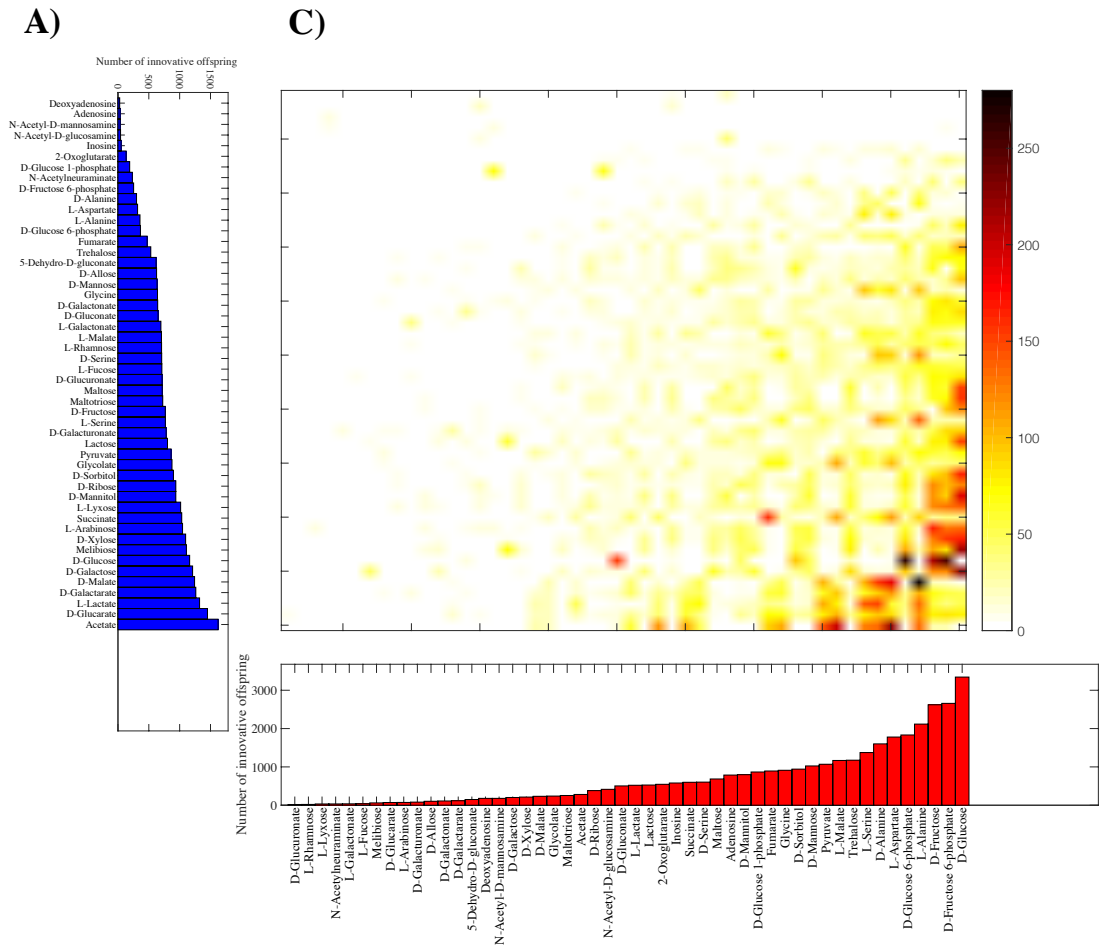


Figure S7: Recombination can create all 50 carbon-use phenotypes considered here ($\|G\|$ = 1,600). A) The horizontal axis shows the number of innovative recombinant offspring (out of one million offspring) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis. This number varies by a factor 58, ranging from 28 on deoxyadenosine to 1,623 on acetate. **B)** Number of innovative recombinants (per million offspring) gaining viability on the novel carbon source specified on the *x*-axis. This number varies by a factor 176, ranging from 19 on D-glucuronate to 3,344 on D-glucose. **C)** Number of innovative recombinants (per million offspring, color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified in panel A, which have gained viability on the novel carbon source specified in panel B. In these analyses, parental metabolic networks contain $\|G\|$ = 1,600 reactions and differ in D = 100 reactions. Moreover, n = 10 reactions are swapped between parental metabolic networks in a recombination event.

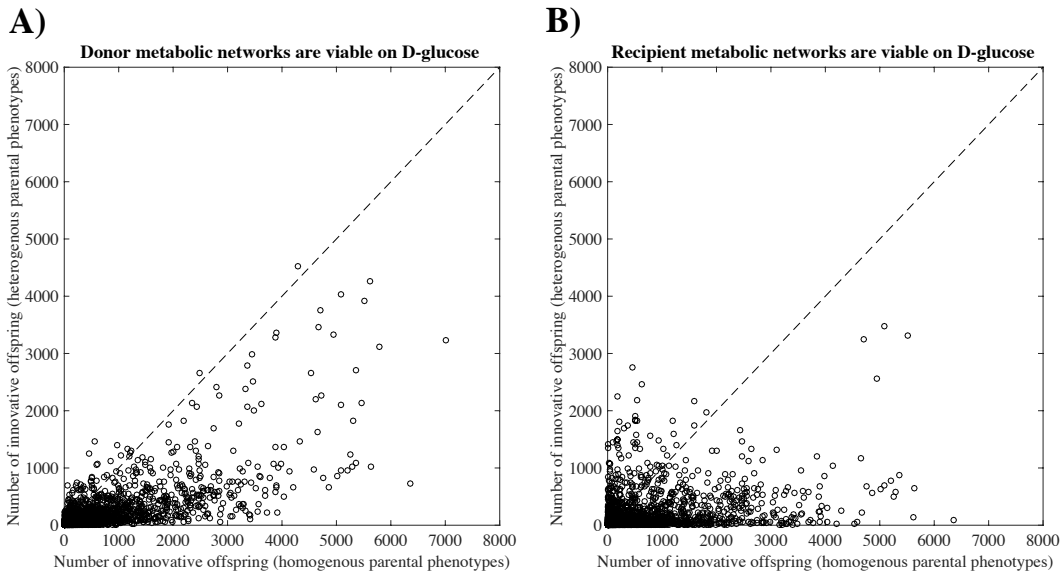


Figure S8: Fewer innovative offspring from phenotypically heterogeneous parents than from phenotypically homogenous parents. Each circle corresponds to a given pair of carbon sources (C_i, C_j) and shows the number of innovative offspring gaining viability on C_j, that are generated by recombination between parents viable on carbon source C_i. The horizontal axis specifies the number of innovative offspring for parents that are viable on the same carbon sources (phenotypically homogeneous parents). The vertical axes show the number of innovative offspring for **A)** parental donors viable on D-glucose and parental recipients viable on C_i, and **B)** parental recipients are viable on D-glucose, and parental donors viable on C_i. In these analyses, all parents have $\|G\|$ = 2,079 reactions, the same as the *E.coli* metabolic network, and their genotypic distance (D) is constant and equals 100. Note that in both panels, the majority of circles (with few exceptions) are placed below the identity ($y = x$) line, indicating that it is more likely for phenotypically homogenous parents to generate innovative offspring than for phenotypically heterogeneous parents.

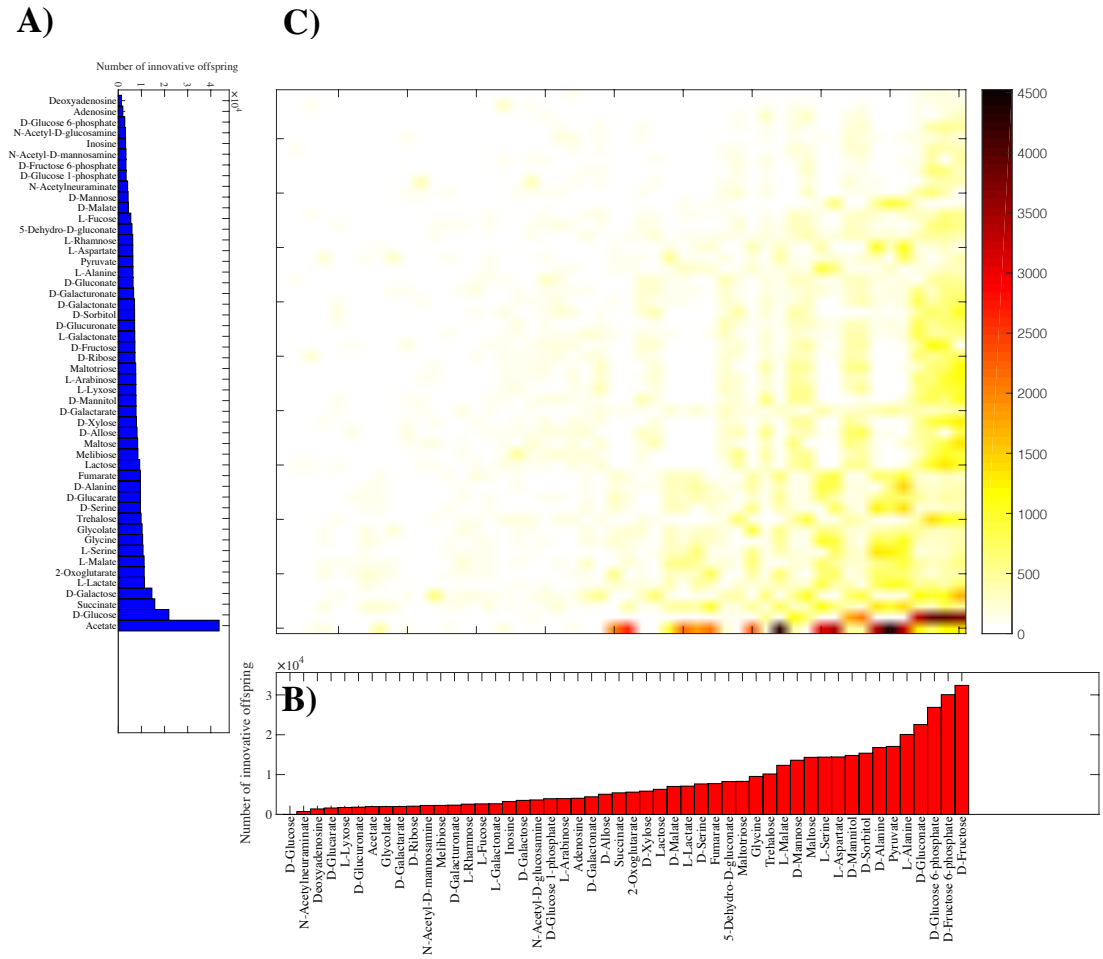


Figure S9: Recombination can create all 50 carbon-use phenotypes considered here (Parents with heterogeneous phenotypes, donors viable only on glucose). A) The horizontal axis shows the number of innovative recombinant offspring (out of one million offspring) resulting from recombination between donor parents viable on glucose and recipient parents that are viable exclusively on the carbon source specified on the vertical axis. This number varies by a factor 32, ranging from 1,371 on deoxyadenosine to 43,615 on acetate. **B)** Number of innovative recombinants (per million offspring) gaining viability on the novel carbon source specified on the *x*-axis. This number varies by a factor 44, ranging from 729 on N-acetylneuraminate to 32,378 on D-fructose. **C)** Number of innovative recombinants (per million offspring, color-coded according to the legend) resulting from recombination between donor parents viable on glucose, and recipient parents viable exclusively on the carbon source specified in panel A, which have gained viability on the novel carbon source specified in panel B. In these analyses, parental metabolic networks contain $\|G\| = 2{,}079$ reactions, the same number as in the *E.coli* metabolic network, and they differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

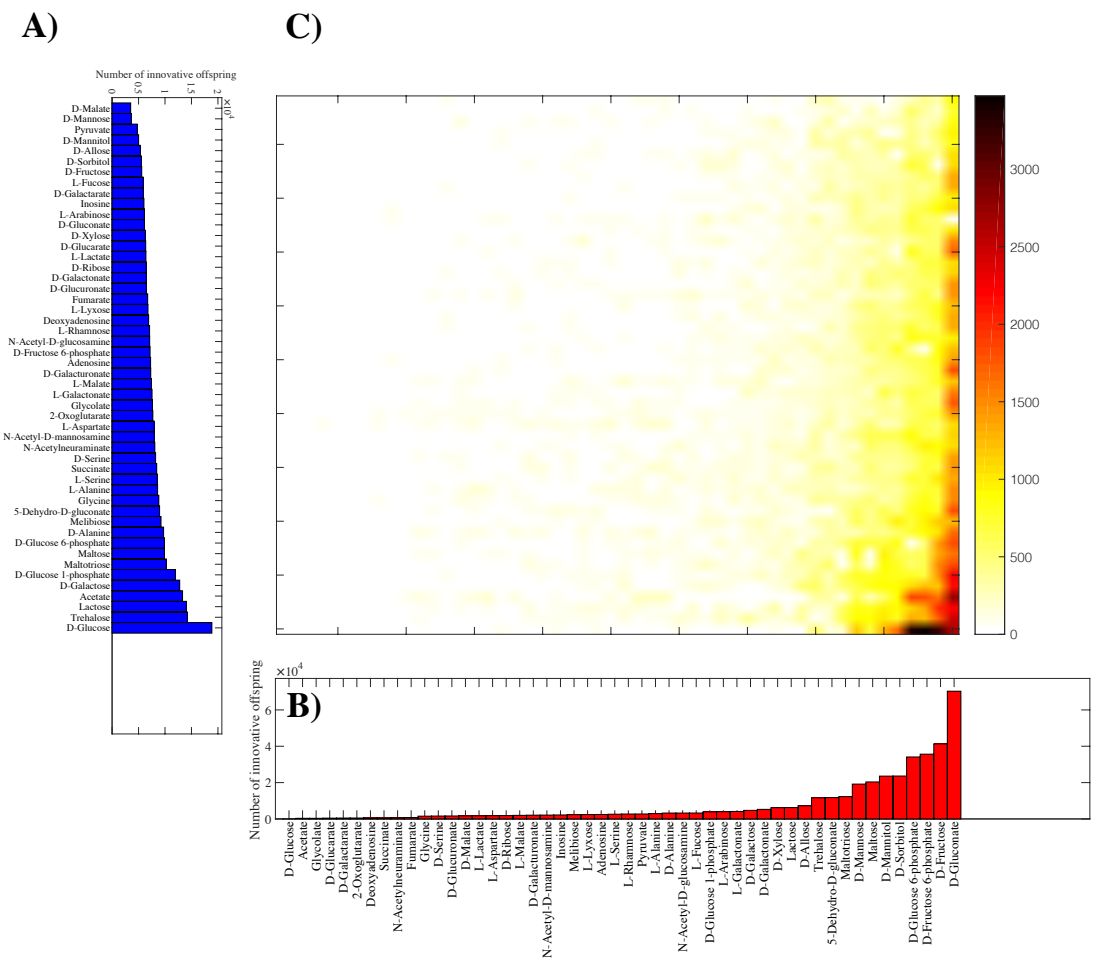
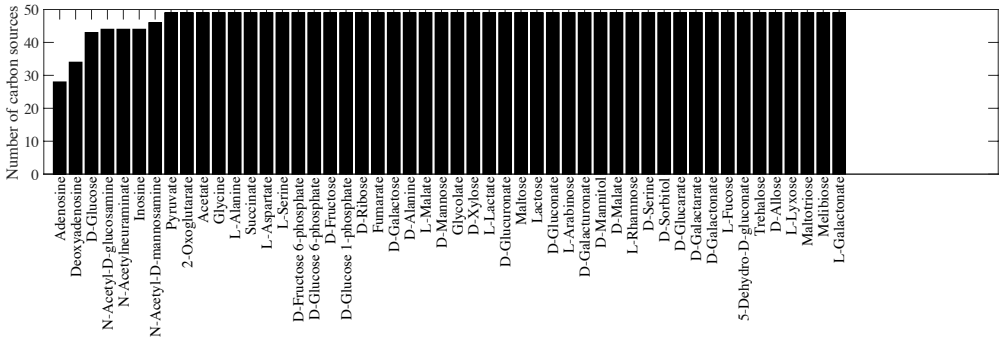
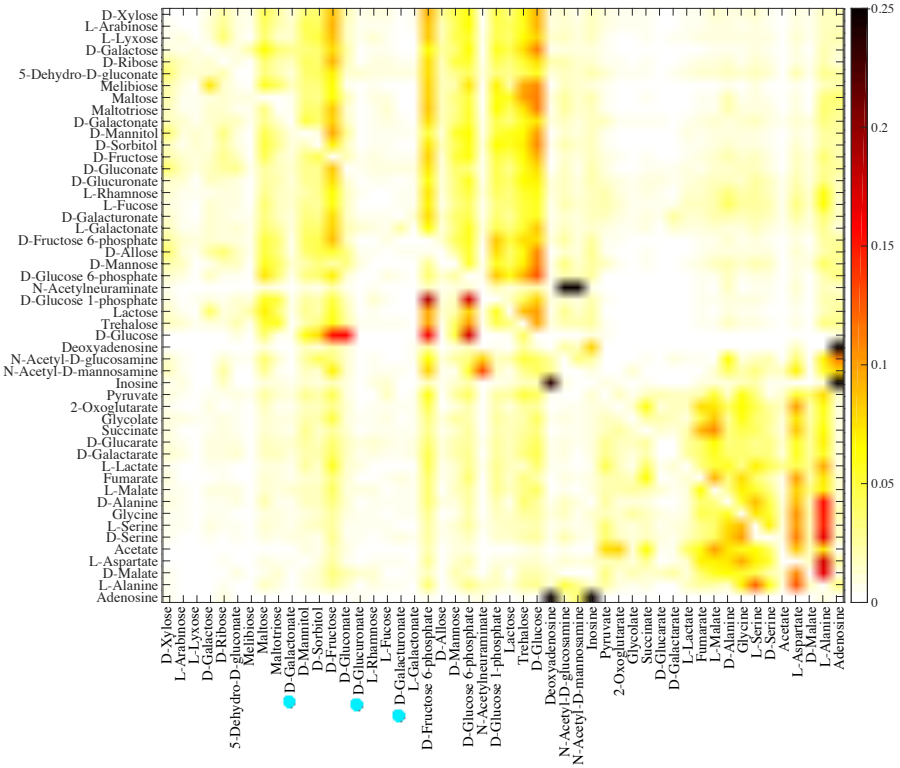


Figure S10: Recombination can create all 50 carbon-use phenotypes considered here (Parents with heterogeneous phenotypes, recipients viable only on glucose). A) The horizontal axis shows the number of innovative recombinant offspring (out of one million offspring) resulting from recombination between recipient parents viable on glucose and donor parents viable exclusively on the carbon source specified on the vertical axis. This number varies by a factor 5, ranging from 3,511 on D-malate to 18,856 on D-glucose. **B)** Number of innovative recombinants (per million offspring) gaining viability on the novel carbon source specified on the *x*-axis. This number varies by a factor 204, ranging from 343 on acetate to 70,292 on D-gluconate. **C)** Number of innovative recombinants (per million offspring, color-coded according to the legend) resulting from recombination between recipient parents viable on glucose, and donor parents viable exclusively on the carbon source specified in panel A, which have gained viability on the novel carbon source specified in panel B. In these analyses, parental metabolic networks contain $\|G\| = 2{,}079$ reactions, the same as in the *E.coli* metabolic network, and they differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Adenosine
Deoxyadenosine
D-Glucose
N-Acetyl-D-glucosamine
N-Acetylneuraminate
Inosine
N-Acetyl-D-mannosamine
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Xylose
L-Arabinose
L-Lyxose
D-Galactose
D-Ribose
5-Dehydro-D-gluconate
Melibiose
Maltose
Maltotriose
D-Galactonate
D-Mannitol
D-Sorbitol
D-Fructose
D-Gluconate
D-Glucuronate
L-Rhamnose
L-Fucose
D-Galacturonate
L-Galactonate
D-Fructose 6-phosphate
D-Allose
D-Mannose
D-Glucose 6-phosphate
N-Acetylneuraminate
D-Glucose 1-phosphate
Lactose
Trehalose
D-Glucose
Deoxyadenosine
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
Inosine
Pyruvate
2-Oxoglutarate
Glycolate
Succinate
D-Glucarate
D-Galactarate
L-Lactate
Fumarate
L-Malate
D-Alanine
Glycine
L-Serine
D-Serine
Acetate
L-Aspartate
D-Malate
L-Alanine
Adenosine
0.25
0.2
0.15
0.1
0.05
0

C)

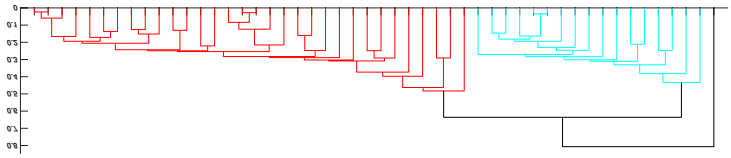
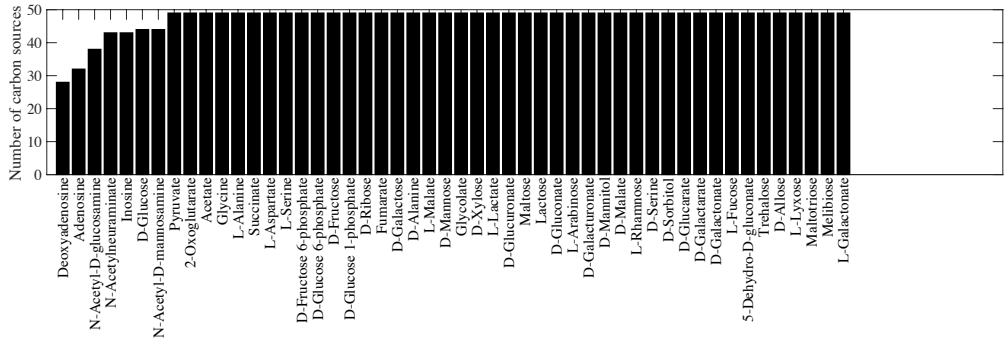
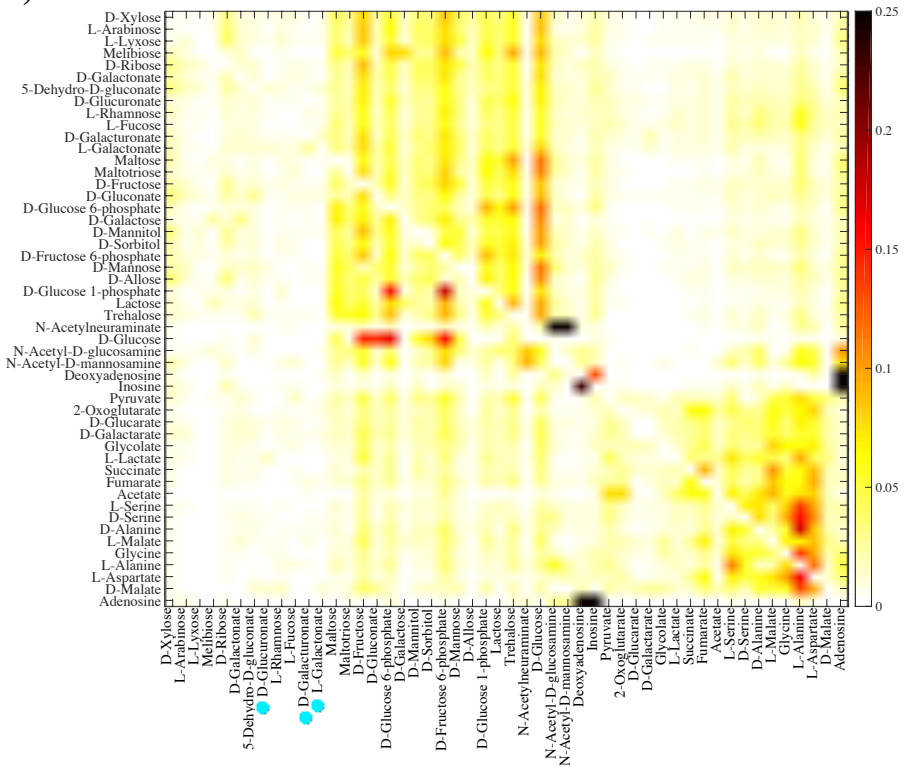


Figure S11: Emergence of innovative offspring can be constrained by parental phenotypes (n = 20). A) The horizontal axis shows the carbon source on which parental metabolisms are viable, and the vertical axis shows the number of novel carbon sources (among the remaining 49 carbon sources) on which at least one innovative offspring results from recombination between parental metabolic networks. **B)** Fraction of innovative recombinants (color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis, which have gained viability on the novel carbon source specified on the horizontal axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate, (shown by cyan circles), which are gluconeogenic carbon sources.). In these analyses, parental metabolic networks contain $\|G\|$ = 2,079 reactions, the same number as in the *E.coli* metabolic network, and they differ in D = 100 reactions. Moreover, n = 20 reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Deoxyadenosine
Adenosine
N-Acetyl-D-glucosamine
N-Acetylneuraminate
Inosine
D-Glucose
N-Acetyl-D-mannosamine
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Xylose
L-Arabinose
L-Lyxose
Melibiose
D-Ribose
D-Galactonate
5-Dehydro-D-gluconate
D-Glucuronate
L-Rhamnose
L-Fucose
D-Galacturonate
L-Galactonate
Maltose
Maltotriose
D-Fructose
D-Gluconate
D-Glucose 6-phosphate
D-Galactose
D-Mannitol
D-Sorbitol
D-Fructose 6-phosphate
D-Mannose
D-Allose
D-Glucose 1-phosphate
Lactose
Trehalose
N-Acetylneuraminate
D-Glucose
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
Deoxyadenosine
Inosine
Pyruvate
2-Oxoglutarate
D-Glucarate
D-Galactarate
Glycolate
L-Lactate
Succinate
Fumarate
Acetate
L-Serine
D-Serine
D-Alanine
L-Malate
Glycine
L-Alanine
L-Aspartate
D-Malate
Adenosine
0.25
0.2
0.15
0.1
0.05
0

C)

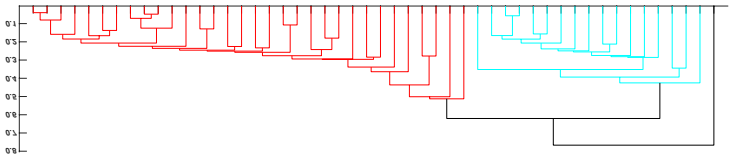
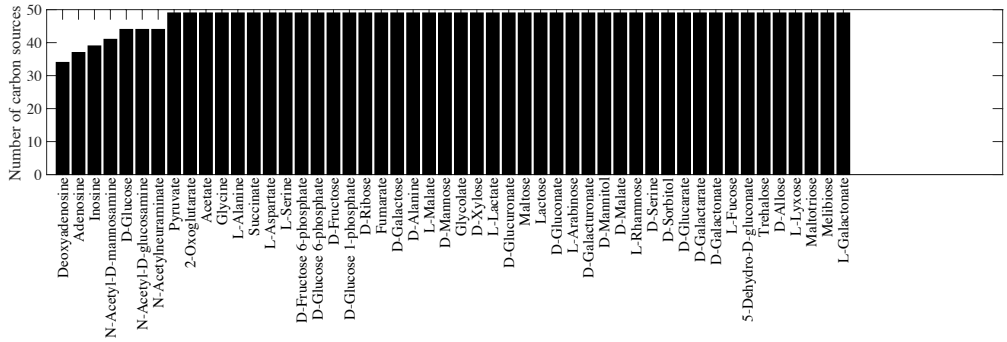
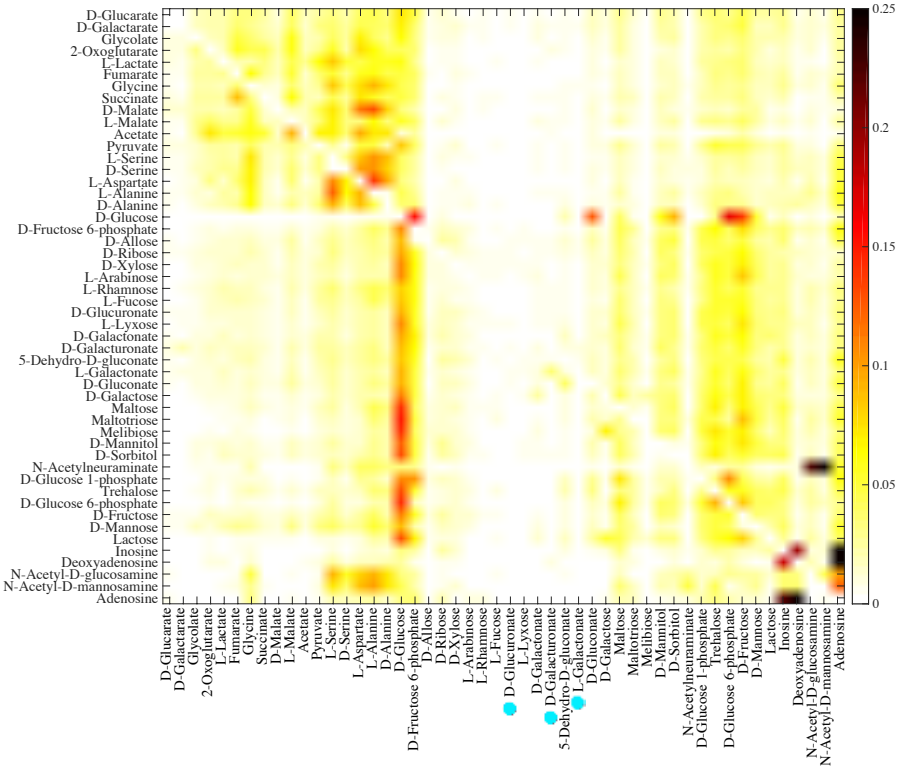


Figure S12: Emergence of innovative offspring can be constrained by parental phenotypes (n = 30). A) The horizontal axis shows the carbon source on which parental metabolisms are viable, and the vertical axis shows the number of novel carbon sources (among the remaining 49 carbon sources) on which at least one innovative offspring results from recombination between parental metabolic networks. **B)** Fraction of innovative recombinants (color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis, which have gained viability on the novel carbon source specified on the horizontal axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources.). In these analyses, parental metabolic networks contain $\|G\|$ = 2,079 reactions, the same number as in the *E.coli* metabolic network, and they differ in D = 100 reactions. Moreover, n = 30 reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
50
40
30
20
10
0
Deoxyadenosine
Adenosine
Inosine
N-Acetyl-D-mannosamine
D-Glucose
N-Acetyl-D-glucosamine
N-Acetylneuraminate
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

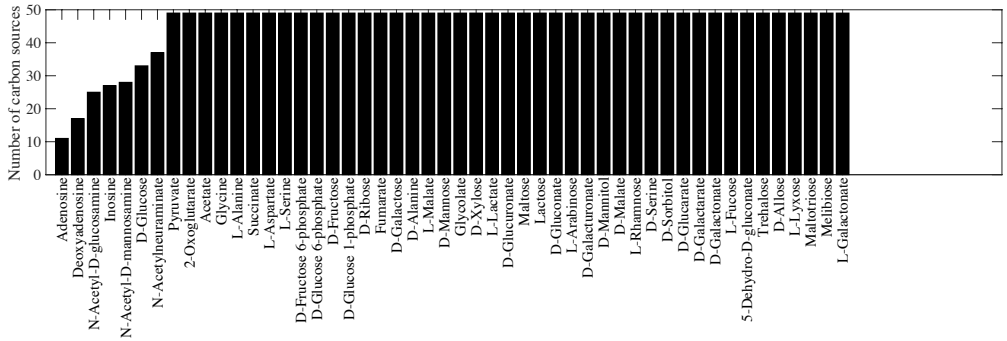
B)


D-Glucarate
D-Galactarate
Glycolate
2-Oxoglutarate
L-Lactate
Fumarate
Glycine
Succinate
D-Malate
L-Malate
Acetate
Pyruvate
L-Serine
D-Serine
L-Aspartate
L-Alanine
D-Alanine
D-Glucose
D-Fructose 6-phosphate
D-Allose
D-Ribose
D-Xylose
L-Arabinose
L-Rhamnose
L-Fucose
D-Glucuronate
L-Lyxose
D-Galactonate
D-Galacturonate
5-Dehydro-D-gluconate
L-Galactonate
D-Gluconate
D-Galactose
Maltose
Maltotriose
Melibiose
D-Mannitol
D-Sorbitol
N-Acetylneuraminate
D-Glucose 1-phosphate
Trehalose
D-Glucose 6-phosphate
D-Fructose
D-Mannose
Lactose
Inosine
Deoxyadenosine
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
Adenosine
0.25
0.2
0.15
0.1
0.05
0

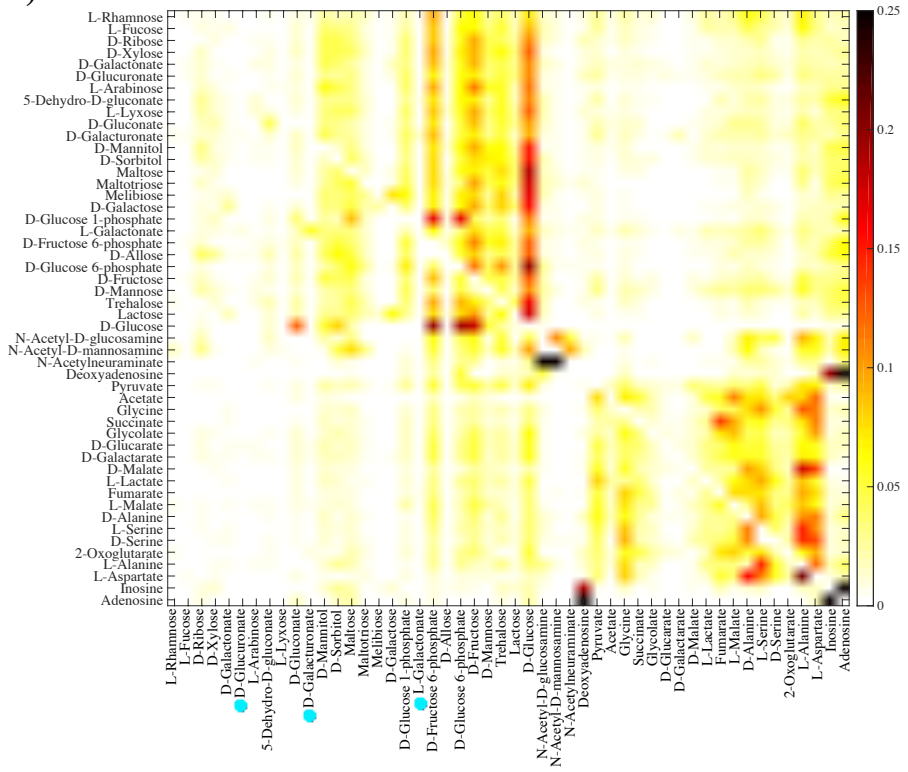
C)

Figure S13: Emergence of innovative offspring can be constrained by parental phenotypes (D = 1,000). A) The horizontal axis shows the carbon source on which parental metabolisms are viable, and the vertical axis shows the number of novel carbon sources (among the remaining 49 carbon sources) on which at least one innovative offspring results from recombination between parental metabolic networks. **B)** Fraction of innovative recombinants (color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis, which have gained viability on the novel carbon source specified on the horizontal axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources.). In these analyses, parental metabolic networks contain $\|G\|$ = 2,079 reactions, the same number as in the *E.coli* metabolic network, and they differ in D = 1,000 reactions. Moreover, n = 10 reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Adenosine
Deoxyadenosine
N-Acetyl-D-glucosamine
Inosine
N-Acetyl-D-mannosamine
D-Glucose
N-Acetylneuraminate
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


L-Rhamnose
L-Fucose
D-Ribose
D-Xylose
D-Galactonate
D-Glucuronate
L-Arabinose
5-Dehydro-D-gluconate
L-Lyxose
D-Gluconate
D-Galacturonate
D-Mannitol
D-Sorbitol
Maltose
Maltotriose
Melibiose
D-Galactose
D-Glucose 1-phosphate
L-Galactonate
D-Fructose 6-phosphate
D-Allose
D-Glucose 6-phosphate
D-Fructose
D-Mannose
Trehalose
Lactose
D-Glucose
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
N-Acetylneuraminate
Deoxyadenosine
Pyruvate
Acetate
Glycine
Succinate
Glycolate
D-Glucarate
D-Galactarate
D-Malate
L-Lactate
Fumarate
L-Malate
D-Alanine
L-Serine
D-Serine
2-Oxoglutarate
L-Alanine
L-Aspartate
Inosine
Adenosine
0.25
0.2
0.15
0.1
0.05
0

C)

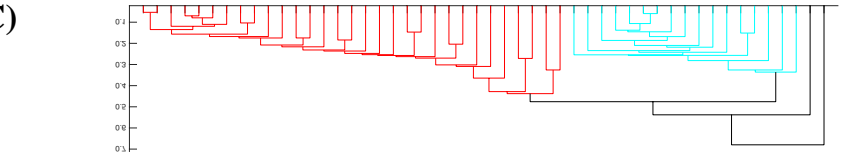
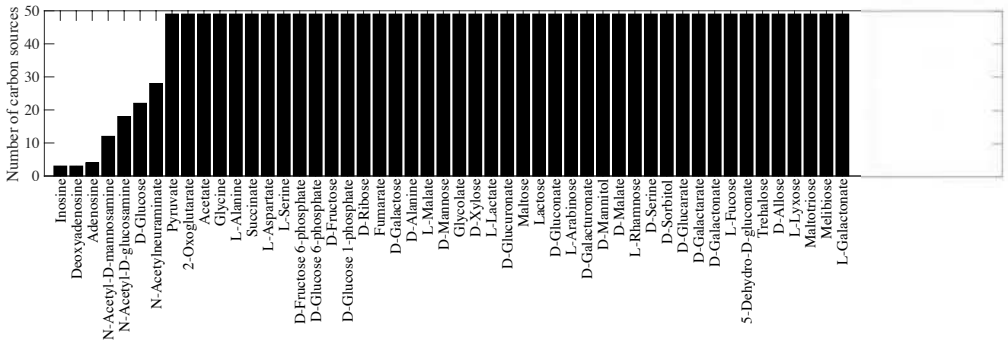
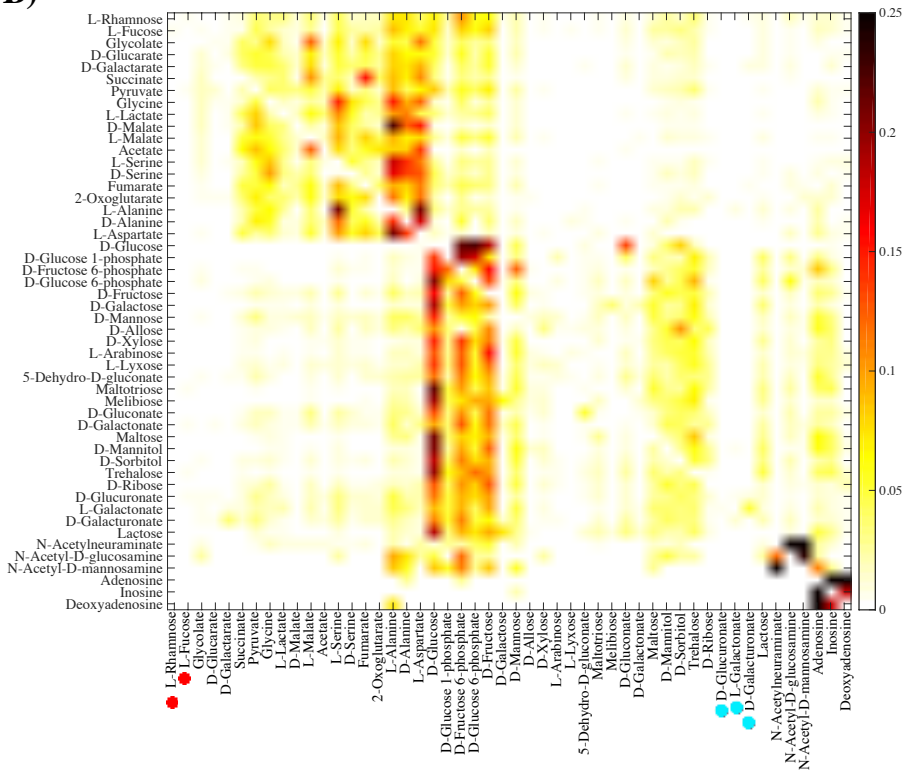


Figure S14: Emergence of innovative offspring can be constrained by parental phenotypes (‖*G*‖ = 1,800). A) The horizontal axis shows the carbon source on which parental metabolisms are viable, and the vertical axis shows the number of novel carbon sources (among the remaining 49 carbon sources) on which at least one innovative offspring results from recombination between parental metabolic networks. **B)** Fraction of innovative recombinants (color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis, which have gained viability on the novel carbon source specified on the horizontal axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources.). In these analyses, parental metabolic networks contain ‖*G*‖ = 1,800 reactions, and they differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Inosine
Deoxyadenosine
Adenosine
N-Acetyl-D-mannosamine
N-Acetyl-D-glucosamine
D-Glucose
N-Acetylneuraminate
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


L-Rhamnose
L-Fucose
Glycolate
D-Glucarate
D-Galactarate
Succinate
Pyruvate
Glycine
L-Lactate
D-Malate
L-Malate
Acetate
L-Serine
D-Serine
Fumarate
2-Oxoglutarate
L-Alanine
D-Alanine
L-Aspartate
D-Glucose
D-Glucose 1-phosphate
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Galactose
D-Mannose
D-Allose
D-Xylose
L-Arabinose
L-Lyxose
5-Dehydro-D-gluconate
Maltotriose
Melibiose
D-Gluconate
D-Galactonate
Maltose
D-Mannitol
D-Sorbitol
Trehalose
D-Ribose
D-Glucuronate
L-Galactonate
D-Galacturonate
Lactose
N-Acetylneuraminate
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
Adenosine
Inosine
Deoxyadenosine
0.25
0.2
0.15
0.1
0.05
0

C)

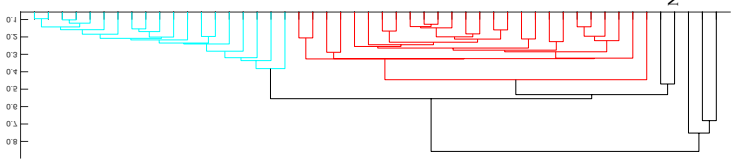
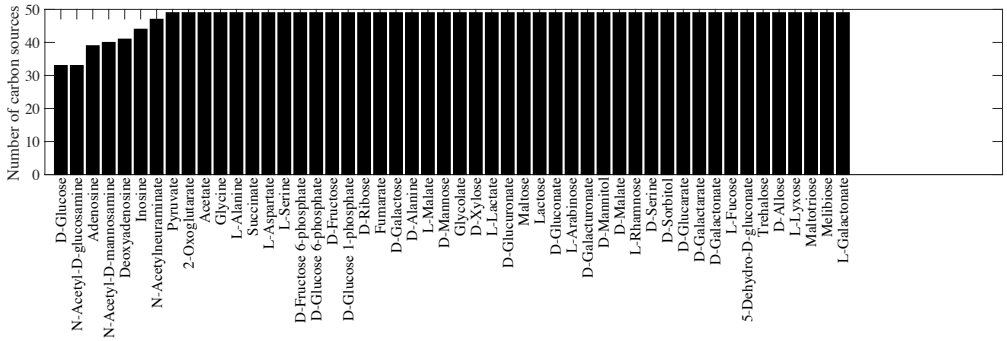
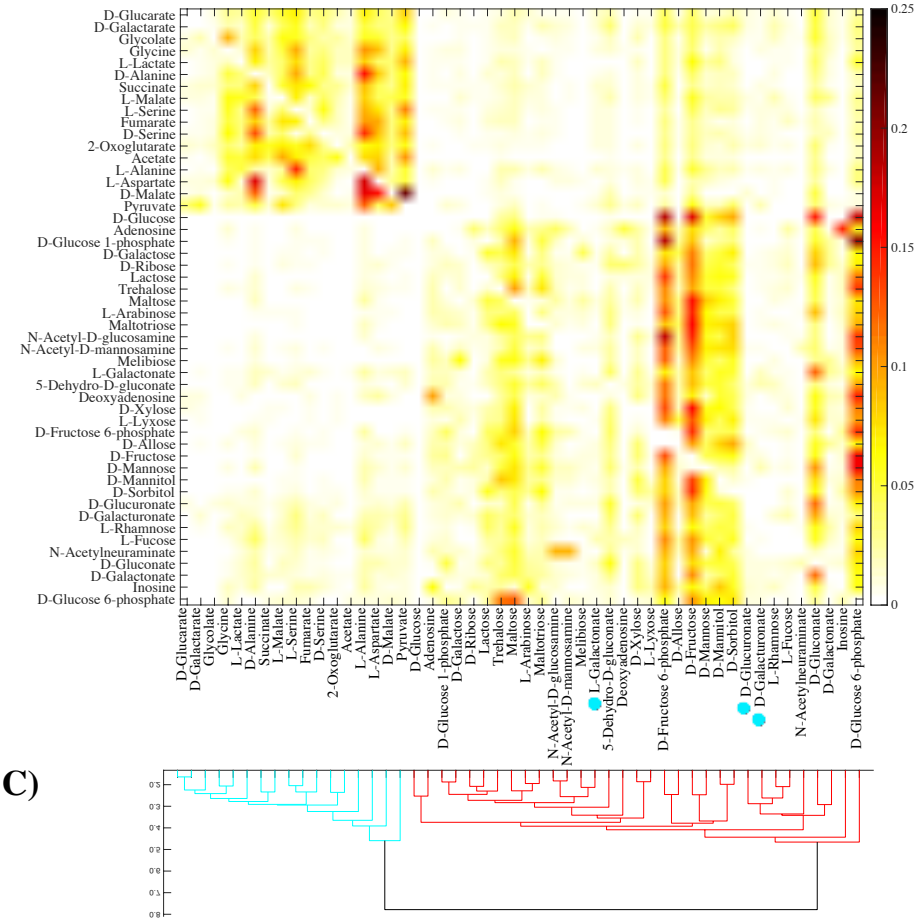


Figure S15: Emergence of innovative offspring can be constrained by parental phenotypes ($\|G\| = 1{,}600$). A) The horizontal axis shows the carbon source on which parental metabolisms are viable, and the vertical axis shows the number of novel carbon sources (among the remaining 49 carbon sources) on which at least one innovative offspring results from recombination between parental metabolic networks. **B)** Fraction of innovative recombinants (color-coded according to the legend) resulting from recombination between parents viable exclusively on the carbon source specified on the vertical axis, which have gained viability on the novel carbon source specified on the horizontal axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources, and L-rhamnose, and L-fucose (shown by red circles), which are glycolytic carbon sources). In these analyses, parental metabolic networks contain $\|G\| = 1{,}600$ reactions, and they differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
D-Glucose
N-Acetyl-D-glucosamine
Adenosine
N-Acetyl-D-mannosamine
Deoxyadenosine
Inosine
N-Acetylneuraminate
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Glucarate
D-Galactarate
Glycolate
Glycine
L-Lactate
D-Alanine
Succinate
L-Malate
L-Serine
Fumarate
D-Serine
2-Oxoglutarate
Acetate
L-Alanine
L-Aspartate
D-Malate
Pyruvate
D-Glucose
Adenosine
D-Glucose 1-phosphate
D-Galactose
D-Ribose
Lactose
Trehalose
Maltose
L-Arabinose
Maltotriose
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
Melibiose
L-Galactonate
5-Dehydro-D-gluconate
Deoxyadenosine
D-Xylose
L-Lyxose
D-Fructose 6-phosphate
D-Allose
D-Fructose
D-Mannose
D-Mannitol
D-Sorbitol
D-Glucuronate
D-Galacturonate
L-Rhamnose
L-Fucose
N-Acetylneuraminate
D-Gluconate
D-Galactonate
Inosine
D-Glucose 6-phosphate
0.25
0.2
0.15
0.1
0.05
0

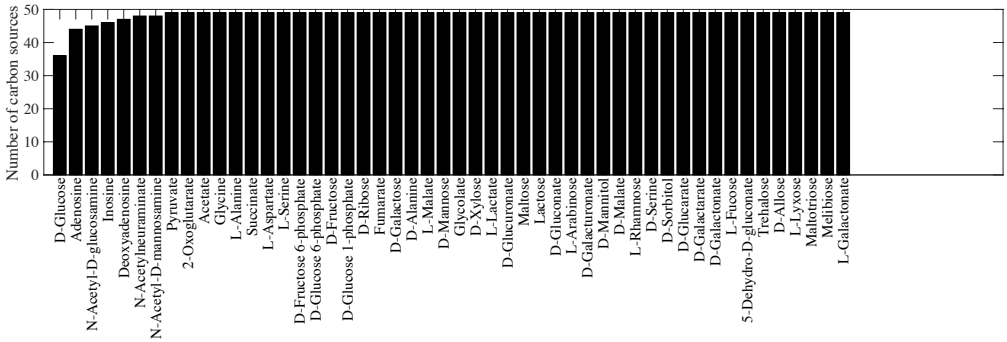
C)

0.5
0.3
0.4
0.2
0.6
0.1
0.8

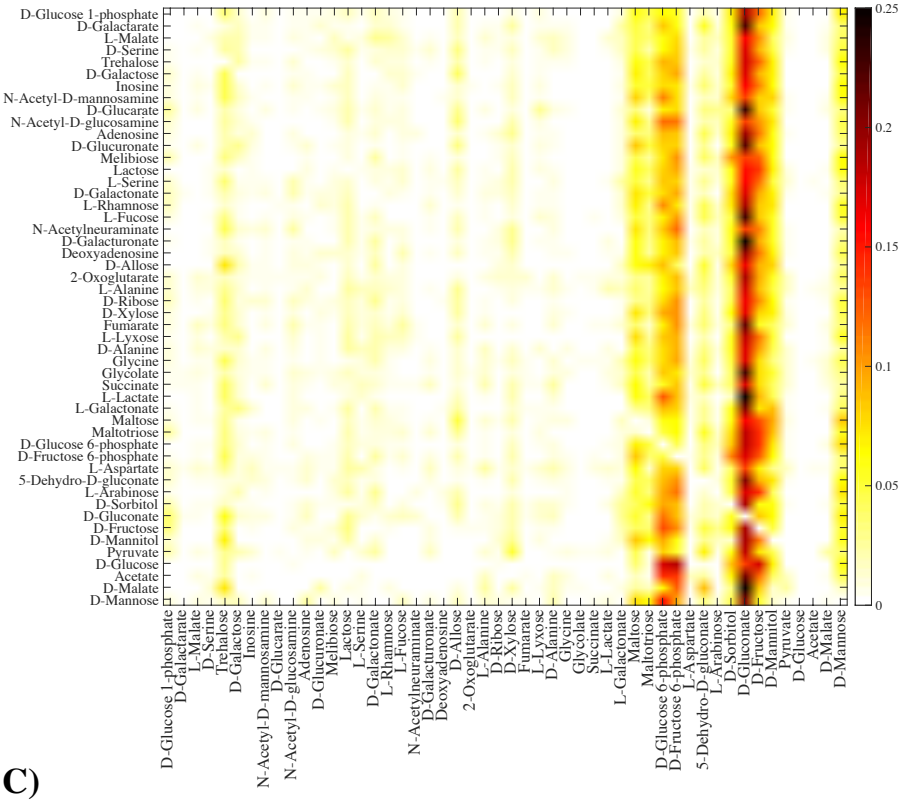
Figure S16: Emergence of innovative offspring can be constrained by parental phenotypes (Parents with heterogeneous phenotypes, donors viable only on glucose). A) The horizontal axis shows the carbon source on which parental metabolisms are viable, and the vertical axis shows the number of novel carbon sources (among the remaining 49 carbon sources) on which at least one innovative offspring results from recombination between parental metabolic networks. **B)** Fraction of innovative recombinants (color-coded according to the legend) resulting from recombination between donor parents viable on glucose and the recipient parents viable exclusively on the carbon source specified on the vertical axis., which have gained viability on the novel carbon source specified on the horizontal axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources.). In these analyses, parental metabolic networks contain $\|G\| = 2{,}079$ reactions, the same number as in the *E.coli* metabolic network, and they differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

=

A)


Number of carbon sources
0
10
20
30
40
50
D-Glucose
Adenosine
N-Acetyl-D-glucosamine
Inosine
Deoxyadenosine
N-Acetylneuraminate
N-Acetyl-D-mannosamine
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Glucose 1-phosphate
D-Galactarate
L-Malate
D-Serine
Trehalose
D-Galactose
Inosine
N-Acetyl-D-mannosamine
D-Glucarate
N-Acetyl-D-glucosamine
Adenosine
D-Glucuronate
Melibiose
Lactose
L-Serine
D-Galactonate
L-Rhamnose
L-Fucose
N-Acetylneuraminate
D-Galacturonate
Deoxyadenosine
D-Allose
2-Oxoglutarate
L-Alanine
D-Ribose
D-Xylose
Fumarate
L-Lyxose
D-Alanine
Glycine
Glycolate
Succinate
L-Lactate
L-Galactonate
Maltose
Maltotriose
D-Glucose 6-phosphate
D-Fructose 6-phosphate
L-Aspartate
5-Dehydro-D-gluconate
L-Arabinose
D-Sorbitol
D-Gluconate
D-Fructose
D-Mannitol
Pyruvate
D-Glucose
Acetate
D-Malate
D-Mannose
0.25
0.2
0.15
0.1
0.05
0

C)

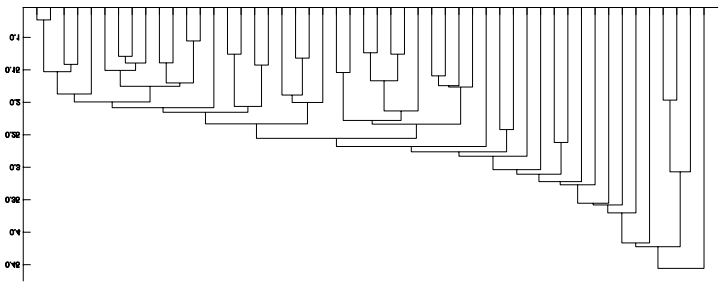
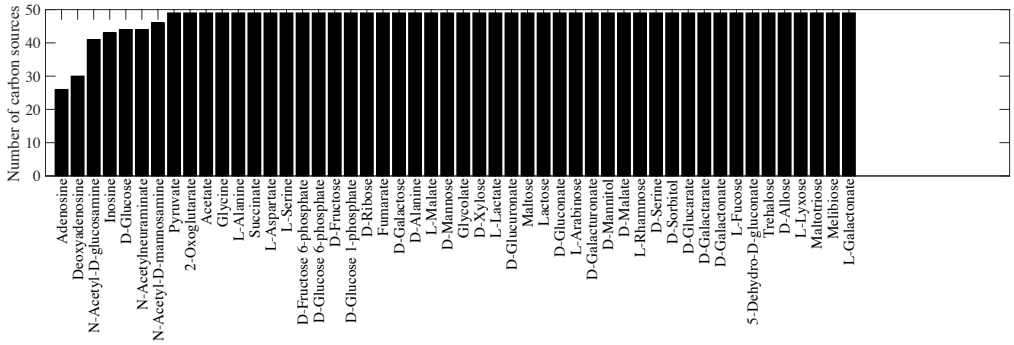
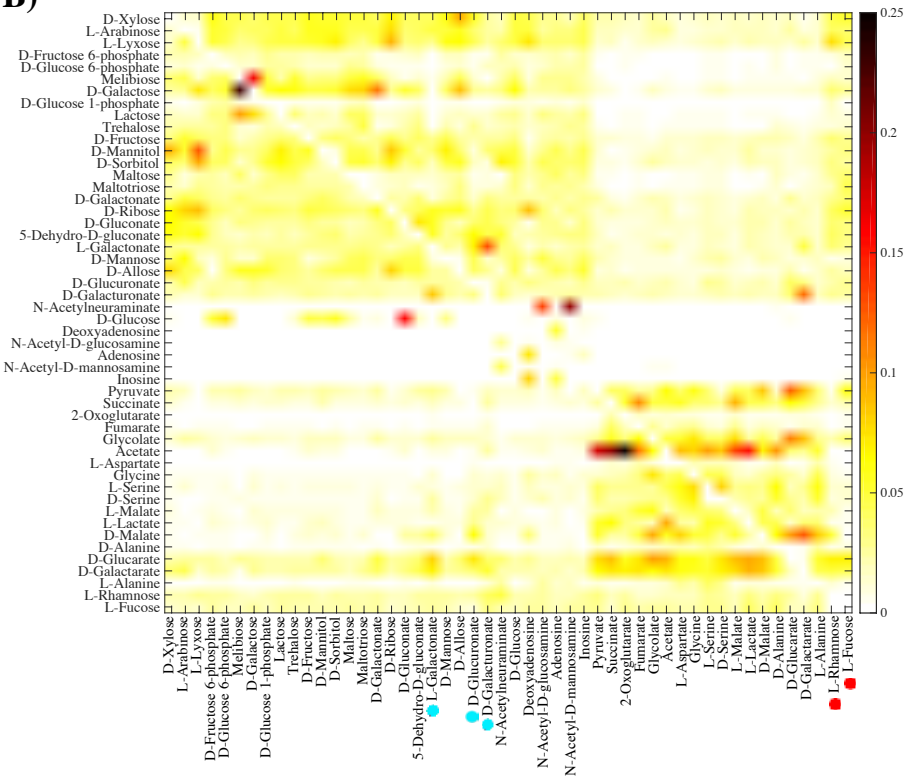


Figure S17: Emergence of innovative offspring can be constrained by parental phenotypes (Parents with heterogeneous phenotypes, recipients viable only on glucose). **A)** The horizontal axis shows the carbon source on which parental metabolisms are viable, and the vertical axis shows the number of novel carbon sources (among the remaining 49 carbon sources) on which at least one innovative offspring results from recombination between parental metabolic networks. **B)** Fraction of innovative recombinants (color-coded according to the legend) resulting from recombination between recipient parents viable on glucose and donor parents viable exclusively on the carbon source specified on the vertical axis., which have gained viability on the novel carbon source specified on the horizontal axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. In this figure, main branches do not reflect glycolytic and gluconeogenic carbon sources as in other figures. In these analyses, parental metabolic networks contain $\|G\| = 2{,}079$ reactions, the same number as in the *E.coli* metabolic network, and they differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Adenosine
Deoxyadenosine
N-Acetyl-D-glucosamine
Inosine
D-Glucose
N-Acetylneuraminate
N-Acetyl-D-mannosamine
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Xylose
L-Arabinose
L-Lyxose
D-Fructose 6-phosphate
D-Glucose 6-phosphate
Melibiose
D-Galactose
D-Glucose 1-phosphate
Lactose
Trehalose
D-Fructose
D-Mannitol
D-Sorbitol
Maltose
Maltotriose
D-Galactonate
D-Ribose
D-Gluconate
5-Dehydro-D-gluconate
L-Galactonate
D-Mannose
D-Allose
D-Glucuronate
D-Galacturonate
N-Acetylneuraminate
D-Glucose
Deoxyadenosine
N-Acetyl-D-glucosamine
Adenosine
N-Acetyl-D-mannosamine
Inosine
Pyruvate
Succinate
2-Oxoglutarate
Fumarate
Glycolate
Acetate
L-Aspartate
Glycine
L-Serine
D-Serine
L-Malate
L-Lactate
D-Malate
D-Alanine
D-Glucarate
D-Galactarate
L-Alanine
L-Rhamnose
L-Fucose
0.25
0.2
0.15
0.1
0.05
0

C)

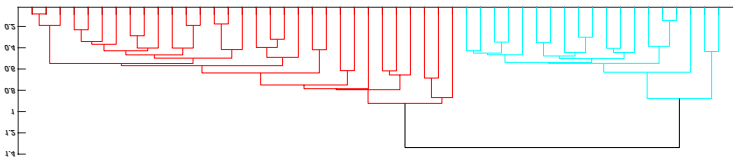
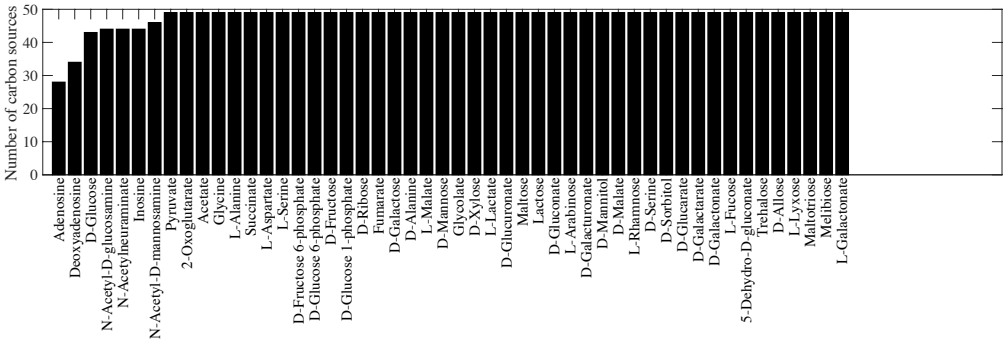
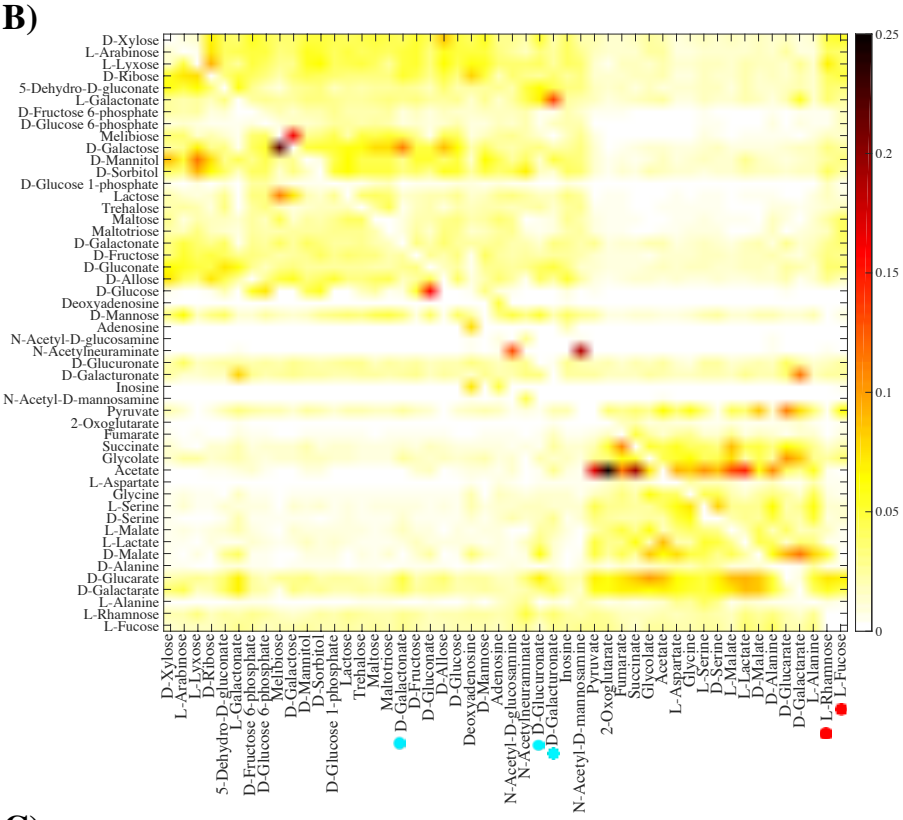


Figure S18: Emergence of innovative offspring is relatively but not absolutely contingent on parental phenotypes. A) The horizontal axis shows the carbon use phenotype C_i of recombinant offspring. The vertical axis shows the number of parental carbon use phenotypes (among 49 possible such phenotypes), from which at least one innovative offspring gained viability on C_i. **B)** Fraction of innovative recombinants (color-coded according to the legend) gaining viability on the novel carbon source specified on the horizontal axis. Recombinants are generated between parents viable exclusively on the carbon source specified on the vertical axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, with the exception of the gluconeogenic carbon sources D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), and the glycolytic carbon sources L-rhamnose, and L-fucose (shown by red circles). In these analyses, parental metabolic networks contain $\|G\|$ = 2,079 reactions, the same number as the *E.coli* metabolic network, and they differ in D = 100 reactions. Moreover, n = 10 reactions are swapped between parental metabolic networks during recombination.

A)


Number of carbon sources
0
10
20
30
40
50
Adenosine
Deoxyadenosine
D-Glucose
N-Acetyl-D-glucosamine
N-Acetylneuraminate
Inosine
N-Acetyl-D-mannosamine
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Xylose
L-Arabinose
L-Lyxose
D-Ribose
5-Dehydro-D-gluconate
L-Galactonate
D-Fructose 6-phosphate
D-Glucose 6-phosphate
Melibiose
D-Galactose
D-Mannitol
D-Sorbitol
D-Glucose 1-phosphate
Lactose
Trehalose
Maltose
Maltotriose
D-Galactonate
D-Fructose
D-Gluconate
D-Allose
D-Glucose
Deoxyadenosine
D-Mannose
Adenosine
N-Acetyl-D-glucosamine
N-Acetylneuraminate
D-Glucuronate
D-Galacturonate
Inosine
N-Acetyl-D-mannosamine
Pyruvate
2-Oxoglutarate
Fumarate
Succinate
Glycolate
Acetate
L-Aspartate
Glycine
L-Serine
D-Serine
L-Malate
L-Lactate
D-Malate
D-Alanine
D-Glucarate
D-Galactarate
L-Alanine
L-Rhamnose
L-Fucose
0.25
0.2
0.15
0.1
0.05
0

C)

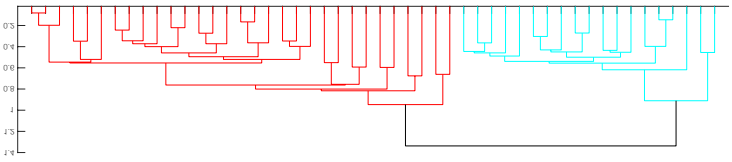
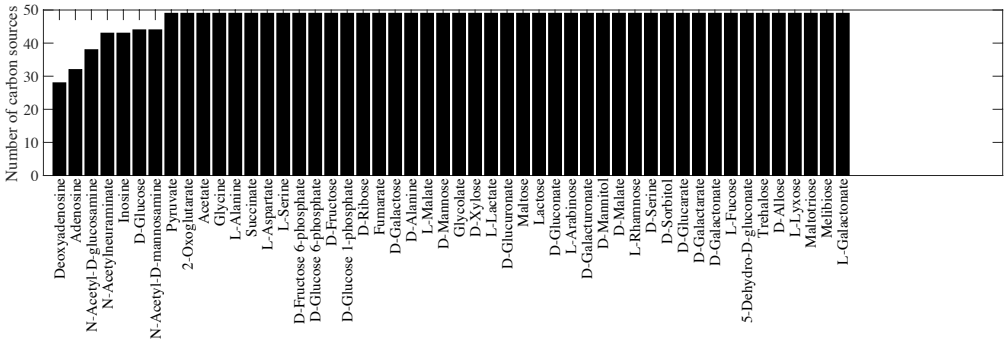
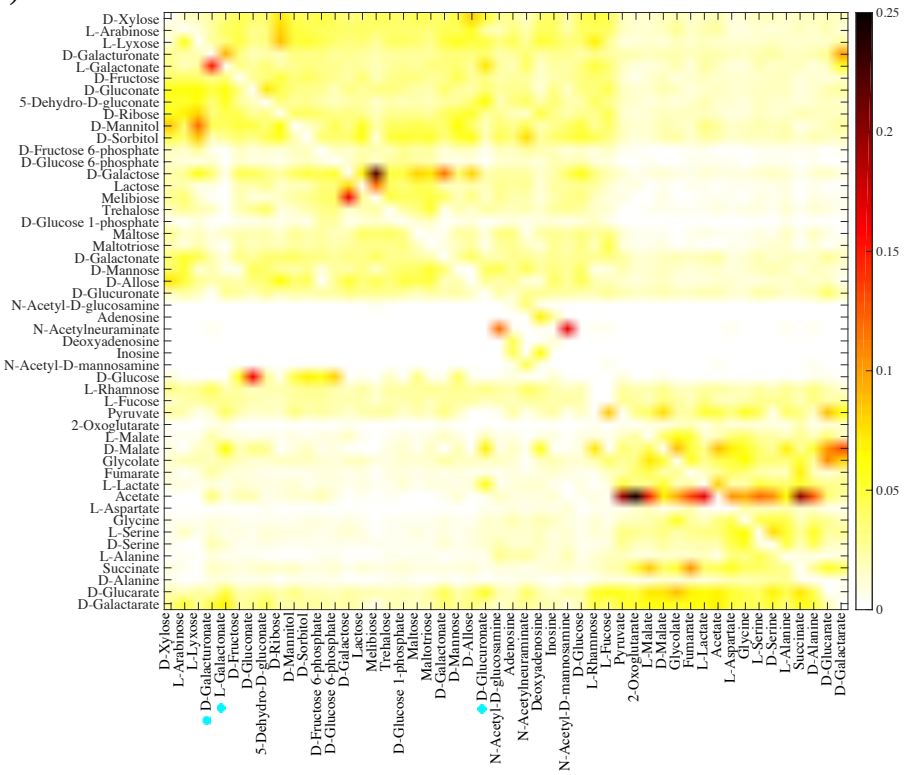


Figure S19: Emergence of innovative offspring is relatively but not absolutely contingent on parental phenotypes (n = 20). A) The horizontal axis shows the carbon use phenotype C_i of recombinant offspring. The vertical axis shows the number of parental carbon use phenotypes (among 49 possible such phenotypes), from which at least one innovative offspring gained viability on C_i. **B)** Fraction of innovative recombinants (color-coded according to the legend) gaining viability on the novel carbon source specified on the horizontal axis, which are generated from recombination between parents viable exclusively on the carbon source specified on the vertical axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources, and L-rhamnose, and L-fucose (shown by red circles), which are glycolytic carbon sources). In these analyses, parental metabolic networks contain $\|G\|$ = 2,079 reactions, the same number as in the *E.coli* metabolic network, and they differ in D = 100 reactions. Moreover, n = 20 reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Deoxyadenosine
Adenosine
N-Acetyl-D-glucosamine
N-AcetylNeuraminate
Inosine
D-Glucose
N-Acetyl-D-mannosamine
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Xylose
L-Arabinose
L-Lyxose
D-Galacturonate
L-Galactonate
D-Fructose
D-Gluconate
5-Dehydro-D-gluconate
D-Ribose
D-Mannitol
D-Sorbitol
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Galactose
Lactose
Melibiose
Trehalose
D-Glucose 1-phosphate
Maltose
Maltotriose
D-Galactonate
D-Mannose
D-Allose
D-Glucuronate
N-Acetyl-D-glucosamine
Adenosine
N-Acetylneuraminate
Deoxyadenosine
Inosine
N-Acetyl-D-mannosamine
D-Glucose
L-Rhamnose
L-Fucose
Pyruvate
2-Oxoglutarate
L-Malate
D-Malate
Glycolate
Fumarate
L-Lactate
Acetate
L-Aspartate
Glycine
L-Serine
D-Serine
L-Alanine
Succinate
D-Alanine
D-Glucarate
D-Galactarate
0.25
0.2
0.15
0.1
0.05
0

C)

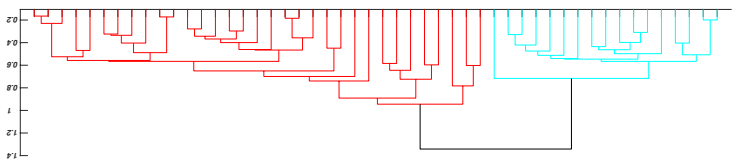
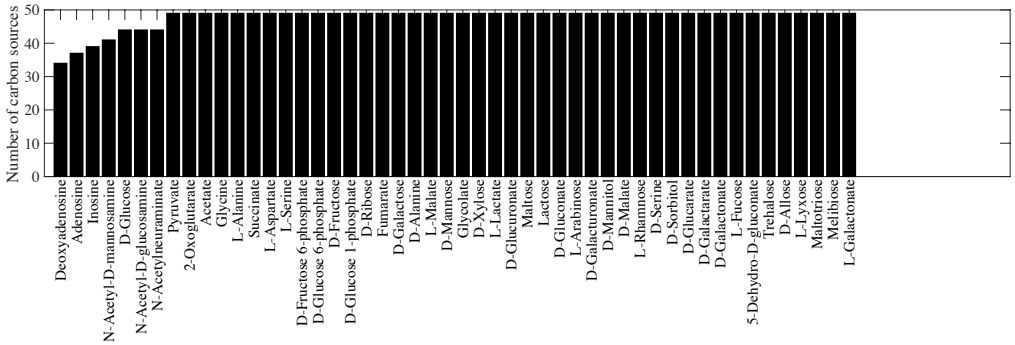
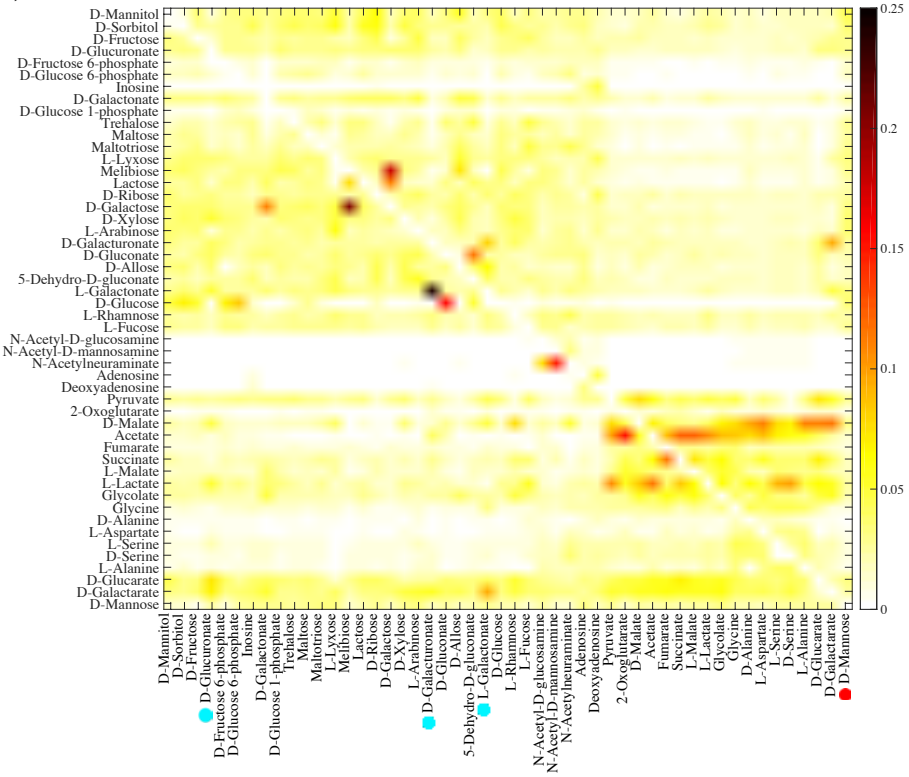


Figure S20: Emergence of innovative offspring is relatively but not absolutely contingent on parental phenotypes (n = 30). A) The horizontal axis shows the carbon use phenotype C_i of recombinant offspring. The vertical axis shows the number of parental carbon use phenotypes (among 49 possible such phenotypes), from which at least one innovative offspring gained viability on C_i. **B)** Fraction of innovative recombinants (color-coded according to the legend) gaining viability on the novel carbon source specified on the horizontal axis, which are generated from recombination between parents viable exclusively on the carbon source specified on the vertical axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources.). In these analyses, parental metabolic networks contain $\|G\|$ = 2,079 reactions, the same number as in the *E.coli* metabolic network, and they differ in D = 100 reactions. Moreover, n = 30 reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Deoxyadenosine
Adenosine
Inosine
N-Acetyl-D-mannosamine
D-Glucose
N-Acetyl-D-glucosamine
N-Acetylneuraminate
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Mannitol
D-Sorbitol
D-Fructose
D-Glucuronate
D-Fructose 6-phosphate
D-Glucose 6-phosphate
Inosine
D-Galactonate
D-Glucose 1-phosphate
Trehalose
Maltose
Maltotriose
L-Lyxose
Melibiose
Lactose
D-Ribose
D-Galactose
D-Xylose
L-Arabinose
D-Galacturonate
D-Gluconate
D-Allose
5-Dehydro-D-gluconate
L-Galactonate
D-Glucose
L-Rhamnose
L-Fucose
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
N-Acetylneuraminate
Adenosine
Deoxyadenosine
Pyruvate
2-Oxoglutarate
D-Malate
Acetate
Fumarate
Succinate
L-Malate
L-Lactate
Glycolate
Glycine
D-Alanine
L-Aspartate
L-Serine
D-Serine
L-Alanine
D-Glucarate
D-Galactarate
D-Mannose
0.25
0.2
0.15
0.1
0.05
0

C)

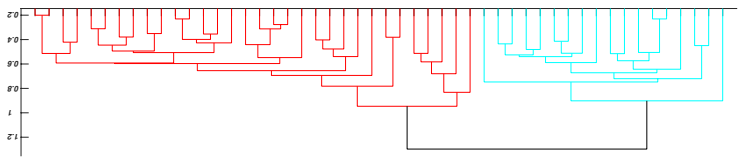
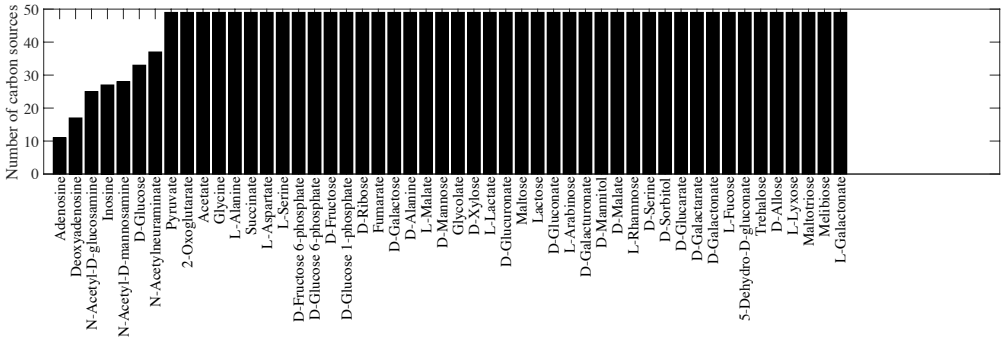
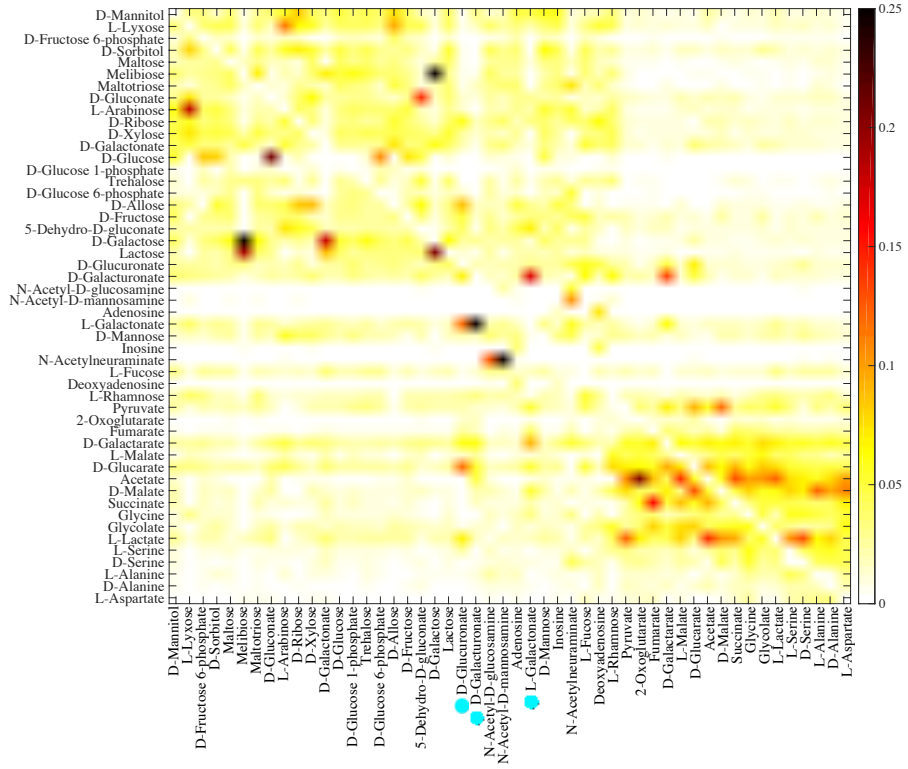


Figure S21: Emergence of innovative offspring is relatively but not absolutely contingent on parental phenotypes (D = 1,000). A) The horizontal axis shows the carbon use phenotype C_i of recombinant offspring. The vertical axis shows the number of parental carbon use phenotypes (among 49 possible such phenotypes), from which at least one innovative offspring gained viability on C_i. **B)** Fraction of innovative recombinants (color-coded according to the legend) gaining viability on the novel carbon source specified on the horizontal axis, which are generated from recombination between parents viable exclusively on the carbon source specified on the vertical axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources, and D-mannose (shown by red circles), which is a glycolytic carbon source). In these analyses, parental metabolic networks contain $\|G\|$ = 2,079 reactions, the same number as in the *E.coli* metabolic network, and they differ in D = 1,000 reactions. Moreover, n = 10 reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Adenosine
Deoxyadenosine
N-Acetyl-D-glucosamine
Inosine
N-Acetyl-D-mannosamine
D-Glucose
N-Acetylneuraminate
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Mannitol
L-Lyxose
D-Fructose 6-phosphate
D-Sorbitol
Maltose
Melibiose
Maltotriose
D-Gluconate
L-Arabinose
D-Ribose
D-Xylose
D-Galactonate
D-Glucose
D-Glucose 1-phosphate
Trehalose
D-Glucose 6-phosphate
D-Allose
D-Fructose
5-Dehydro-D-gluconate
D-Galactose
Lactose
D-Glucuronate
D-Galacturonate
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
Adenosine
L-Galactonate
D-Mannose
Inosine
N-Acetylneuraminate
L-Fucose
Deoxyadenosine
L-Rhamnose
Pyruvate
2-Oxoglutarate
Fumarate
D-Galactarate
L-Malate
D-Glucarate
Acetate
D-Malate
Succinate
Glycine
Glycolate
L-Lactate
L-Serine
D-Serine
L-Alanine
D-Alanine
L-Aspartate
0.25
0.2
0.15
0.1
0.05
0

C)

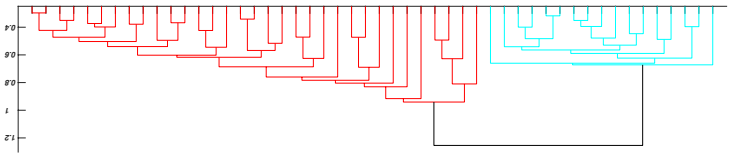
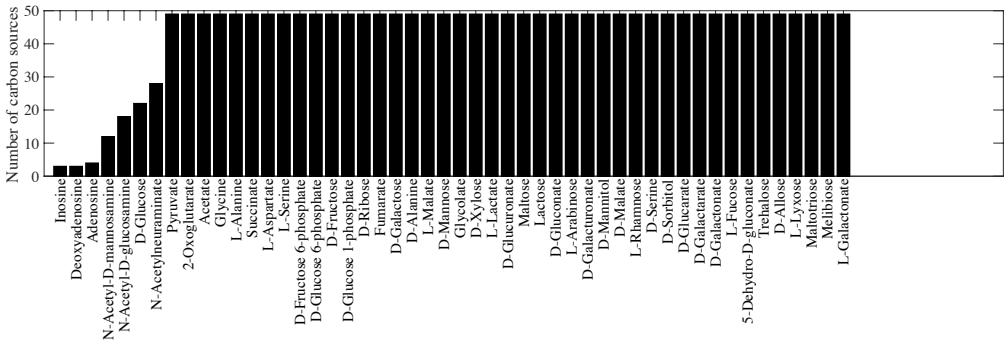
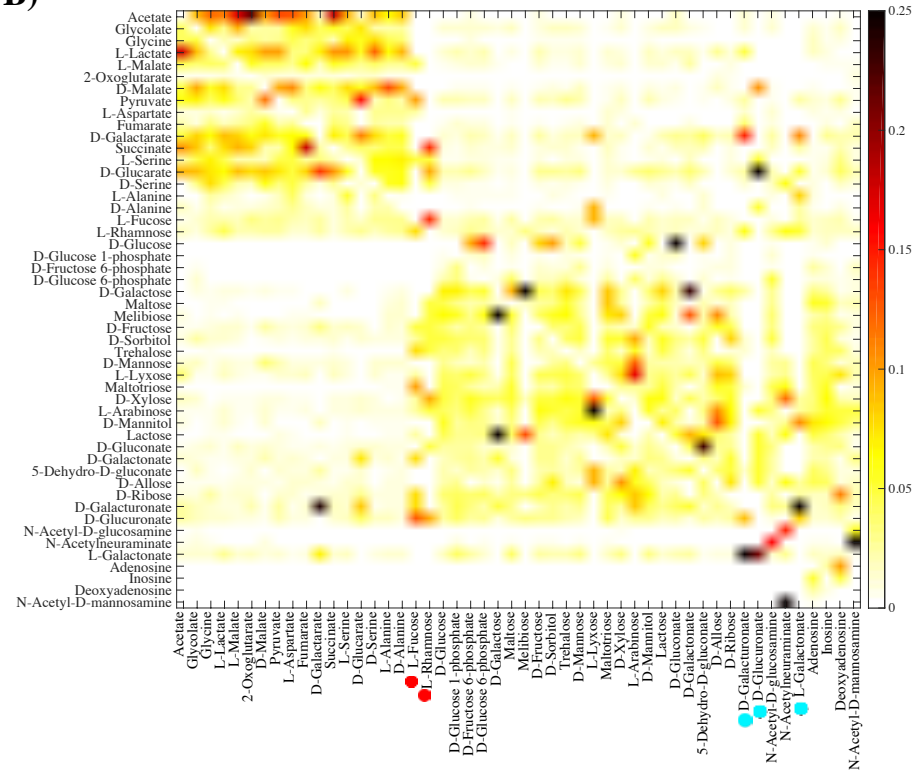

0.4
0.6
0.8
1
1.2

Figure S22: Emergence of innovative offspring is relatively but not absolutely contingent on parental phenotypes ($\|G\| = 1{,}800$). A) The horizontal axis shows the carbon use phenotype C_i of recombinant offspring. The vertical axis shows the number of parental carbon use phenotypes (among 49 possible such phenotypes), from which at least one innovative offspring gained viability on C_i. **B)** Fraction of innovative recombinants (color-coded according to the legend) gaining viability on the novel carbon source specified on the horizontal axis, which are generated from recombination between parents viable exclusively on the carbon source specified on the vertical axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources). In these analyses, parental metabolic networks contain $\|G\| =$ 1,800 reactions, and they differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
Inosine
Deoxyadenosine
Adenosine
N-Acetyl-D-mannosamine
N-Acetyl-D-glucosamine
D-Glucose
N-Acetylneuraminate
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


Acetate
Glycolate
Glycine
L-Lactate
L-Malate
2-Oxoglutarate
D-Malate
Pyruvate
L-Aspartate
Fumarate
D-Galactarate
Succinate
L-Serine
D-Glucarate
D-Serine
L-Alanine
D-Alanine
L-Fucose
L-Rhamnose
D-Glucose
D-Glucose 1-phosphate
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Galactose
Maltose
Melibiose
D-Fructose
D-Sorbitol
Trehalose
D-Mannose
L-Lyxose
Maltotriose
D-Xylose
L-Arabinose
D-Mannitol
Lactose
D-Gluconate
D-Galactonate
5-Dehydro-D-gluconate
D-Allose
D-Ribose
D-Galacturonate
D-Glucuronate
N-Acetyl-D-glucosamine
N-Acetylneuraminate
L-Galactonate
Adenosine
Inosine
Deoxyadenosine
N-Acetyl-D-mannosamine
0.25
0.2
0.15
0.1
0.05
0

C)

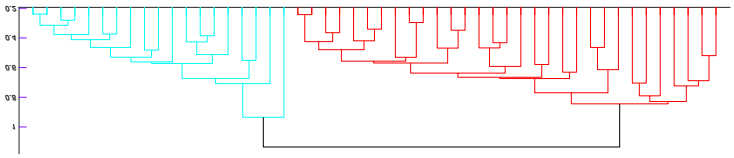
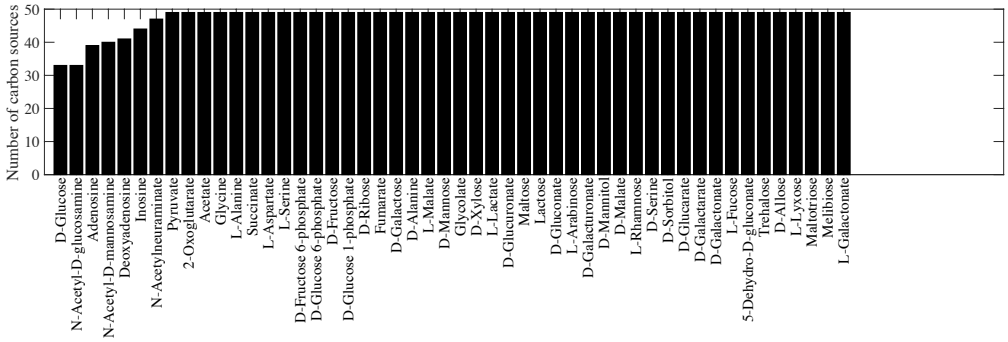
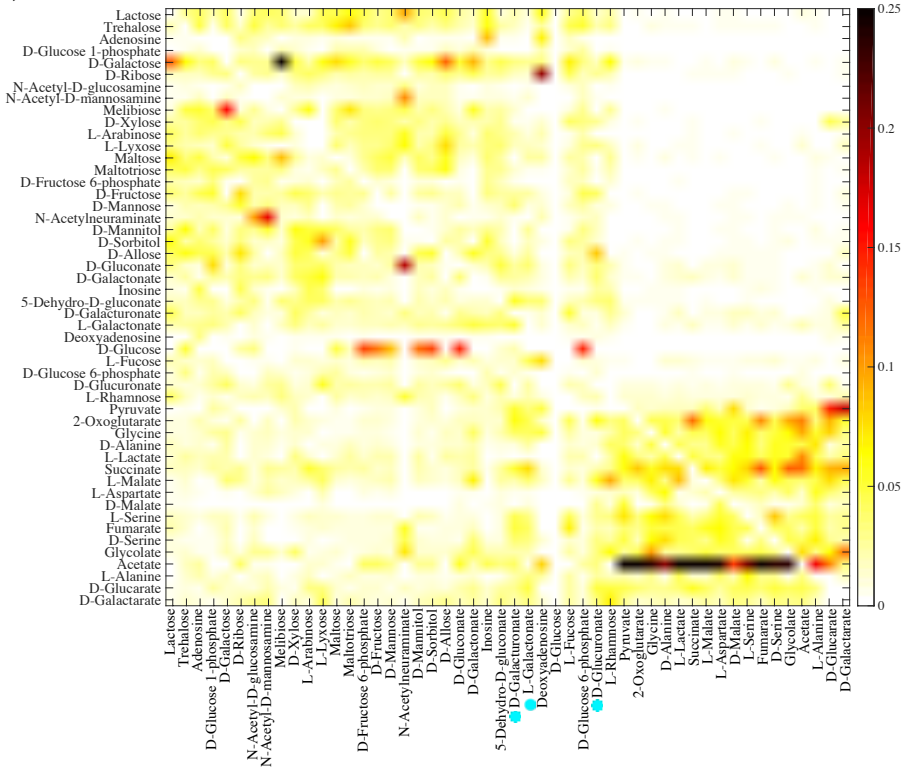


Figure S23: Emergence of innovative offspring is relatively but not absolutely contingent on parental phenotypes ($\|G\| = 1{,}600$). A) The horizontal axis shows the carbon use phenotype C_i of recombinant offspring. The vertical axis shows the number of parental carbon use phenotypes (among 49 possible such phenotypes), from which at least one innovative offspring gained viability on C_i. **B)** Fraction of innovative recombinants (color-coded according to the legend) gaining viability on the novel carbon source specified on the horizontal axis, which are generated from recombination between parents viable exclusively on the carbon source specified on the vertical axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources, and L-rhamnose, and L-fucose (shown by red circles), which are glycolytic carbon sources). . In these analyses, parental metabolic networks contain $\|G\| = 1{,}600$ reactions, and they differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
D-Glucose
N-Acetyl-D-glucosamine
Adenosine
N-Acetyl-D-mannosamine
Deoxyadenosine
Inosine
N-Acetylneuraminate
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


Lactose
Trehalose
Adenosine
D-Glucose 1-phosphate
D-Galactose
D-Ribose
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
Melibiose
D-Xylose
L-Arabinose
L-Lyxose
Maltose
Maltotriose
D-Fructose 6-phosphate
D-Fructose
D-Mannose
N-Acetylneuraminate
D-Mannitol
D-Sorbitol
D-Allose
D-Gluconate
D-Galactonate
Inosine
5-Dehydro-D-gluconate
D-Galacturonate
L-Galactonate
Deoxyadenosine
D-Glucose
L-Fucose
D-Glucose 6-phosphate
D-Glucuronate
L-Rhamnose
Pyruvate
2-Oxoglutarate
Glycine
D-Alanine
L-Lactate
Succinate
L-Malate
L-Aspartate
D-Malate
L-Serine
Fumarate
D-Serine
Glycolate
Acetate
L-Alanine
D-Glucarate
D-Galactarate
0.25
0.2
0.15
0.1
0.05
0

C)

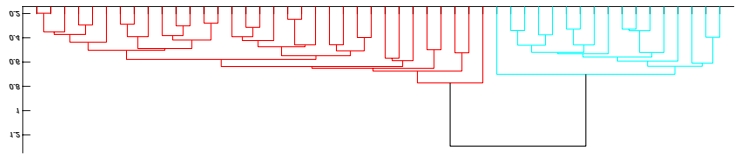
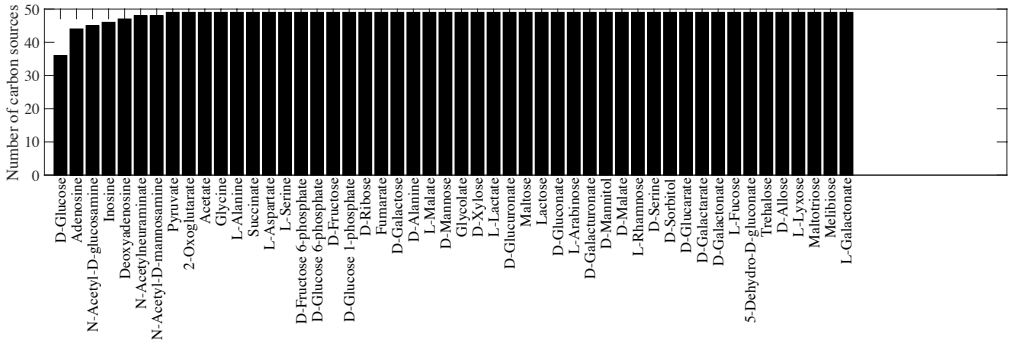
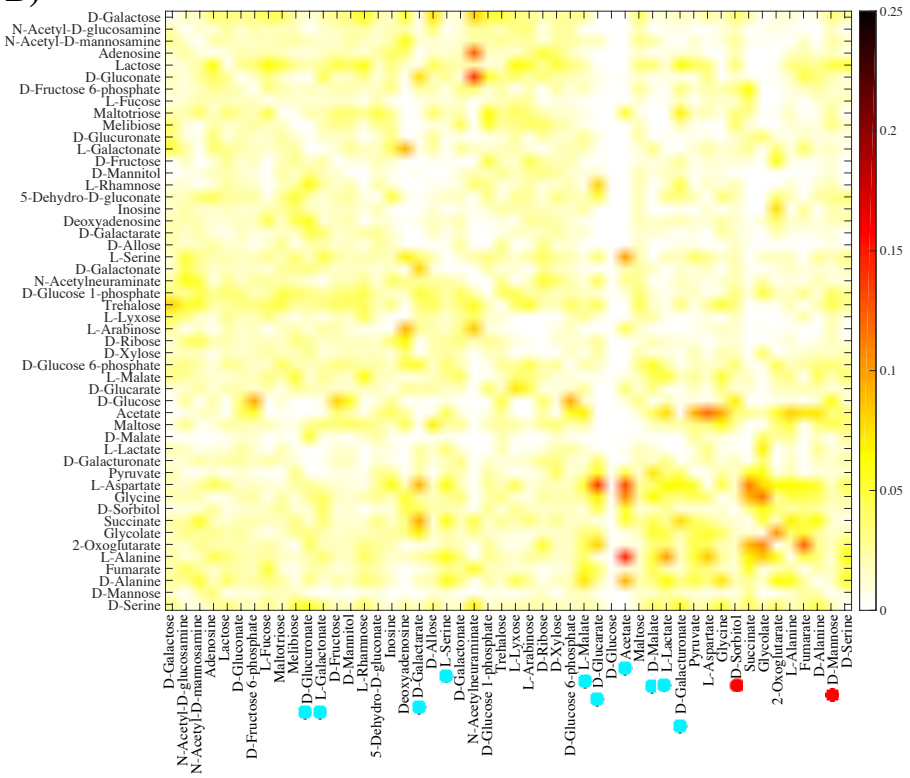


Figure S24: Emergence of innovative offspring is relatively but not absolutely contingent on parental phenotypes (Parents with heterogeneous phenotypes, donors viable only on glucose). A) The horizontal axis shows the carbon use phenotype C_i of recombinant offspring. The vertical axis shows the number of parental carbon use phenotypes (among 49 possible such phenotypes), from which at least one innovative offspring gained viability on C_i. **B)** Fraction of innovative recombinants (color-coded according to the legend) gaining viability on the novel carbon source specified on the horizontal axis, which are generated from recombination between donor parents viable exclusively on glucose and the recipient parents that are exclusively viable on the carbon source specified on the vertical axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (except D-galacturonate, L-galactonate, and D-glucoronate (shown by cyan circles), which are gluconeogenic carbon sources.). In these analyses, parental metabolic networks contain $\|G\| = 1{,}800$ reactions, and differ in $D = 100$ reactions. Moreover, $n = 10$ reactions are swapped between parental metabolic networks in a recombination event.

A)


Number of carbon sources
0
10
20
30
40
50
D-Glucose
Adenosine
N-Acetyl-D-glucosamine
Inosine
Deoxyadenosine
N-Acetylneuraminate
N-Acetyl-D-mannosamine
Pyruvate
2-Oxoglutarate
Acetate
Glycine
L-Alanine
Succinate
L-Aspartate
L-Serine
D-Fructose 6-phosphate
D-Glucose 6-phosphate
D-Fructose
D-Glucose 1-phosphate
D-Ribose
Fumarate
D-Galactose
D-Alanine
L-Malate
D-Mannose
Glycolate
D-Xylose
L-Lactate
D-Glucuronate
Maltose
Lactose
D-Gluconate
L-Arabinose
D-Galacturonate
D-Mannitol
D-Malate
L-Rhamnose
D-Serine
D-Sorbitol
D-Glucarate
D-Galactarate
D-Galactonate
L-Fucose
5-Dehydro-D-gluconate
Trehalose
D-Allose
L-Lyxose
Maltotriose
Melibiose
L-Galactonate

B)


D-Galactose
N-Acetyl-D-glucosamine
N-Acetyl-D-mannosamine
Adenosine
Lactose
D-Gluconate
D-Fructose 6-phosphate
L-Fucose
Maltotriose
Melibiose
D-Glucuronate
L-Galactonate
D-Fructose
D-Mannitol
L-Rhamnose
5-Dehydro-D-gluconate
Inosine
Deoxyadenosine
D-Galactarate
D-Allose
L-Serine
D-Galactonate
N-Acetylneuraminate
D-Glucose 1-phosphate
Trehalose
L-Lyxose
L-Arabinose
D-Ribose
D-Xylose
D-Glucose 6-phosphate
L-Malate
D-Glucarate
D-Glucose
Acetate
Maltose
D-Malate
L-Lactate
D-Galacturonate
Pyruvate
L-Aspartate
Glycine
D-Sorbitol
Succinate
Glycolate
2-Oxoglutarate
L-Alanine
Fumarate
D-Alanine
D-Mannose
D-Serine
0.25
0.2
0.15
0.1
0.05
0

C)

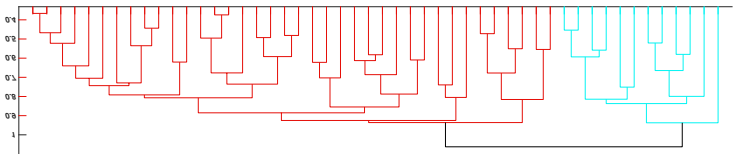


Figure S25: Emergence of innovative offspring is relatively but not absolutely contingent on parental phenotypes (Parents with heterogeneous phenotypes, recipients viable only on glucose). A) The horizontal axis shows the carbon use phenotype C_i of recombinant offspring. The vertical axis shows the number of parental carbon use phenotypes (among 49 possible such phenotypes), from which at least one innovative offspring gained viability on C_i. **B)** Fraction of innovative recombinants (color-coded according to the legend) gaining viability on the novel carbon source specified on the horizontal axis, which are generated from recombination between recipient parents viable exclusively on glucose and donor parents that are exclusively viable on the carbon source specified on the vertical axis. **C)** Dendrogram of carbon sources clustered based on their "innovation distance" defined by the data in panel B. We used UPGMA (unweighted pair group method with arithmetic means) for clustering carbon sources. Branches colored in red (cyan) correspond to glycolytic and gluconeogenic carbon sources, (with 12 exceptions; shown by 10 cyan circles, and 2 red circles.). In these analyses, parental metabolic networks contain $\|G\|$ = 1,800 reactions, and differ in D = 100 reactions. Moreover, n = 10 reactions are swapped between parental metabolic networks in a recombination event.

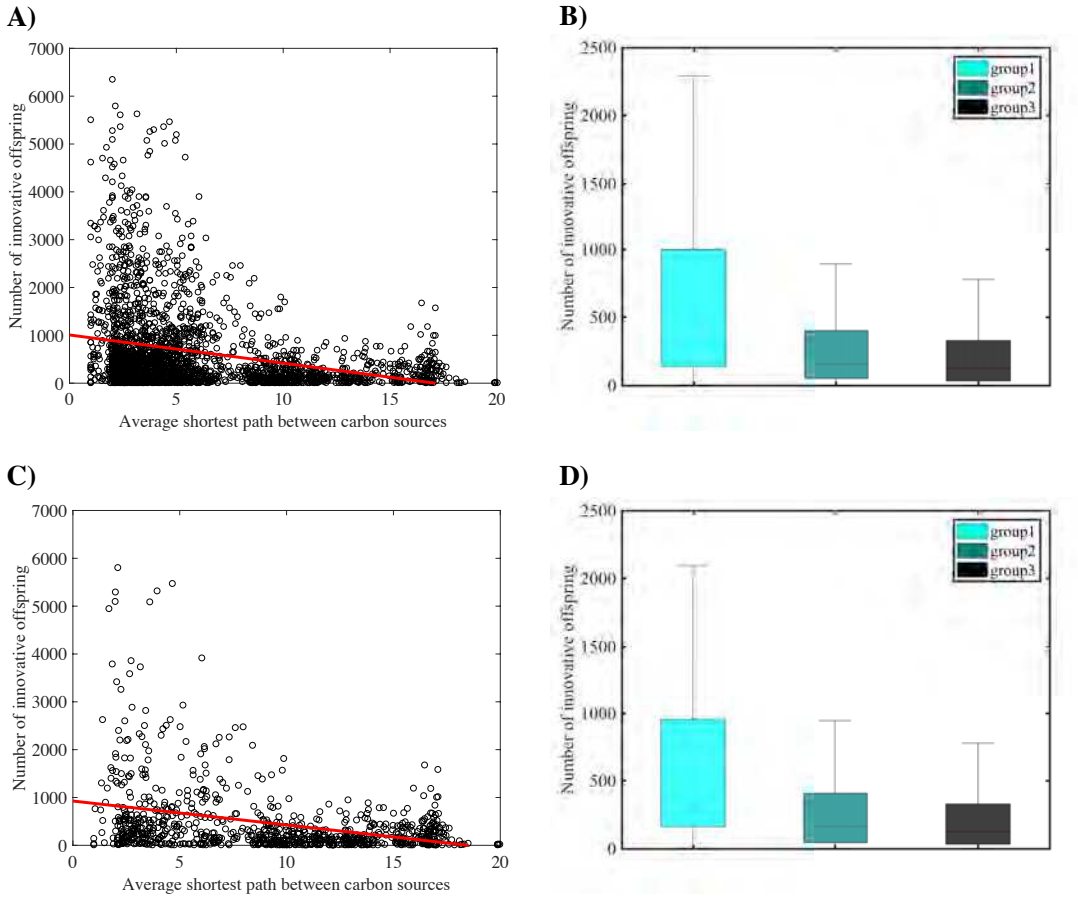


Figure S26: Distance between carbon sources in substrate graphs and relative constraint in the emergence of innovative offspring. In all 4 panels, the vertical axis shows the number of innovative recombinants (per 1 million recombinant offspring) gaining viability on some new carbon source C_j resulting from recombination between parental metabolic networks viable on carbon source C_i. In panels A and C, the horizontal axes show the mean shortest path between carbon source C_i and C_j in the substrate graph (supplementary text S7) of the metabolic networks viable on carbon source C_i. In panel **A)** each circle corresponds to a given pair of carbon sources (C_i, C_j), and data on both axes are significantly correlated (Pearson r=-*0.2722*, and $P<10^{-41}$). In panel **B)** the carbon source pairs (C_i, C_j) are divided into three groups based on their mean shortest path ($||SP\ (i,j)||$) between carbon source C_i and C_j in the substrate graph of metabolic networks viable on carbon source C_i: group 1 $\{i,j|1 \leq ||SP\ (i,j)|| \leq 6\}$), group 2 $\{i,j|6 < ||SP\ (i,j)|| \leq 1\}$), and group 3 $\{i,j|\ ||SP\ (i,j)|| > 12\}$. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima.
In panel A, a non-uniform distribution of mean shortest paths ($||SP\ (i,j)||$) between carbon sources is evident on the horizontal axis. To exclude the possibility that the correlation in panel A is significant simply because of a higher number of data points for lower shortest path distances, we repeated the analyses shown in panels A and B by resampling from the 2500 pairs of carbon sources an equal number of pairs in each distance category, i.e., 284 pairs (C_i,

C_j) with $\{i,j|1 \leq ||SP\ (i,j)|| \leq 6\}$), 284 pairs ($C_i$, C_j) with $\{i,j|6 < ||SP\ (i,j)|| \leq 12\}$), and 284 pairs ($C_i$, C_j) with $\{i,j|\ ||SP\ (i,j)|| > 12\}$, to create the subsampled data in panels C and D. In panel **C)** each circle corresponds to a given pair of carbon sources (C_i, C_j), and data on both axes are significantly correlated (Pearson r=*-0.3411*, and P<10^{-24}). In panel **D)**, analogous to panel B, carbon source pairs (C_i, C_j) are divided into three equally-sized groups based on their mean shortest path ($||SP\ (i,j)||$) between carbon source C_i and C_j in the substrate graph of metabolic networks viable on carbon source C_i: group 1 $\{i,j|1 \leq ||SP\ (i,j)|| \leq 6\}$), group 2 $\{i,j|6 < ||SP\ (i,j)|| \leq 1\}$), and group 3 $\{i,j|\ ||SP\ (i,j)|| > 12\}$. Boxes span the 25-th to 75-th percentile, and whiskers indicate maxima and minima. In these analyses, parental metabolic networks contain ‖G‖=2079 reactions, the same as the *E.coli* metabolic network, and they differ in *D=100* reactions. Moreover, *n=10* reactions are swapped between parental metabolic networks during recombination.

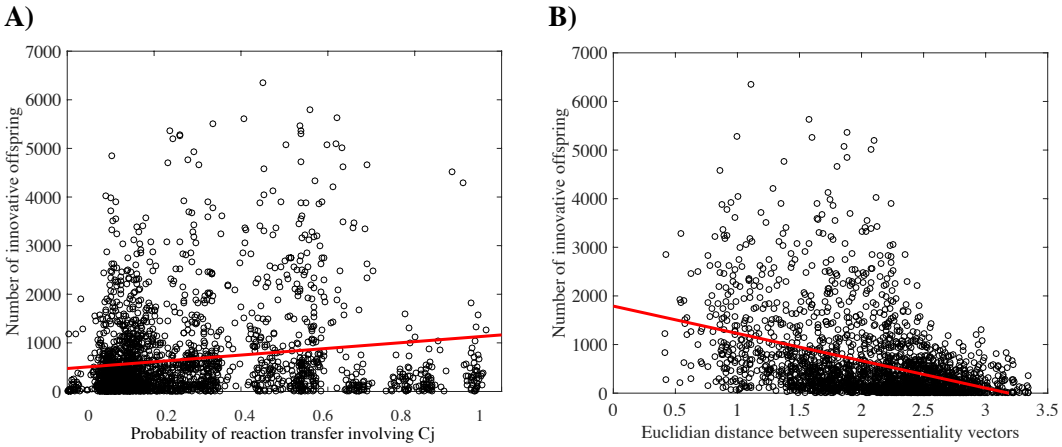


Figure S27: In both panels, each circle corresponds to a given pair of carbon sources (C_i, C_j) and the vertical axis shows the number of innovative recombinants (per 1 million recombinant offspring) gaining viability on some new carbon source C_j resulting from recombination between parental metabolic networks viable on carbon source C_i. The horizontal axes show **A)** the fraction of parental metabolic network pairs viable on carbon source C_i, in which a reaction that can enable viability on carbon source C_j can be transferred from the donor to the recipient metabolic network, and **B)** the Euclidian distance between superessentiality vectors of the corresponding pair of carbon sources, which we use as another proxy for the biochemical distance between carbon sources. In both panels the data plotted against one another are significantly correlated: **A)** Pearson r=*0.163*, and P<10^{-15}, and **B)** Pearson r=*-0.3935*, and P<10^{-83}. In these analyses, parental metabolic networks contain ‖G‖=2079 reactions, the same as the *E.coli* metabolic network, and they differ in *D=100* reactions. Moreover, *n=10* reactions are swapped between parental metabolic networks during recombination.

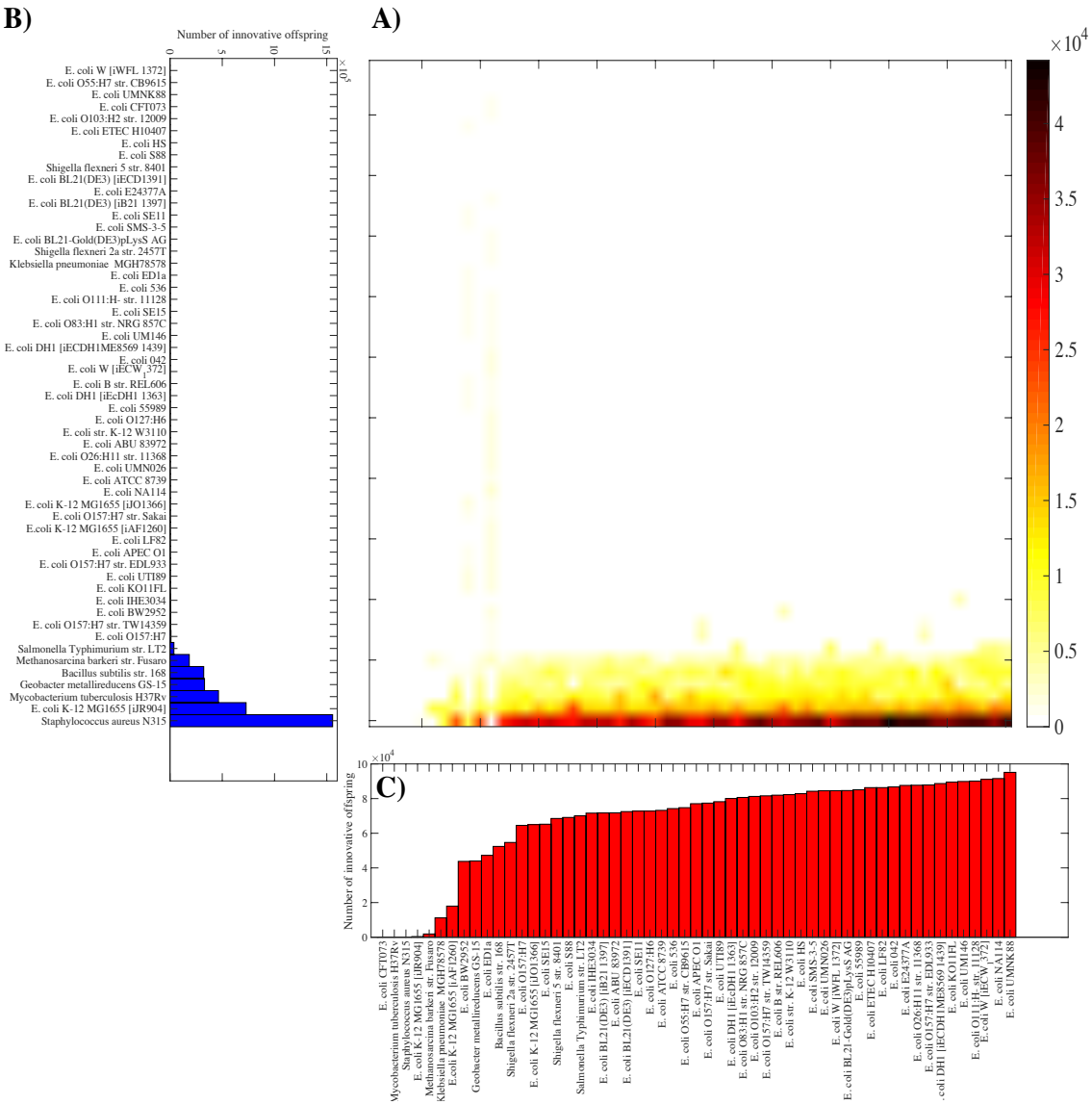


Figure S28: Emergence of innovative offspring is contingent on and constrained by parental genotypes. A) Number of innovative offspring resulting from linkage-based recombination between bacterial DNA donors specified on the vertical axis of panel B, and the corresponding recipient genotypes specified on the horizontal axis of panel C (coded according to the color legend). **B)** Total number of innovative recombinant offspring involving the donor genotype specified on the vertical axis. **C)** Total number of innovative recombinant offspring involving the recipient genotype specified on the horizontal axis.

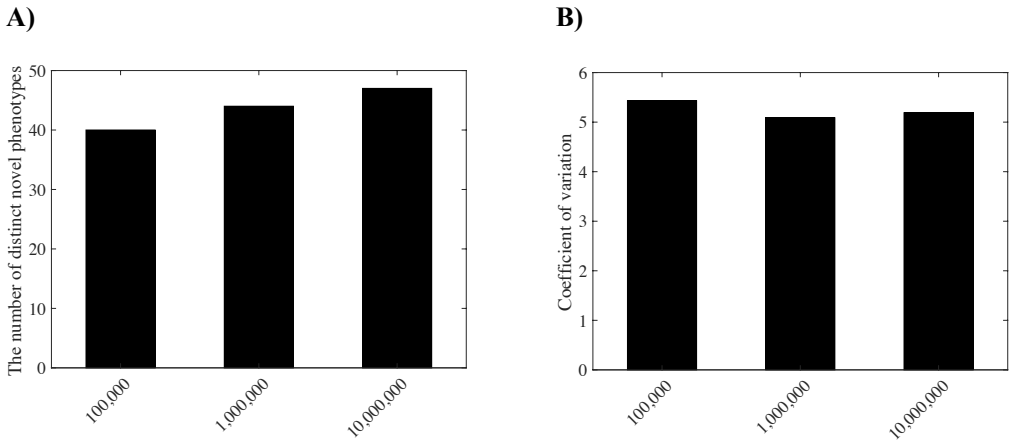


Figure S29: Sample size and its effect on absolute and relative constraints. For this analysis, we used 1,000 parental metabolic networks that are viable exclusively on glucose, and in three different simulations we generated *i)* 100, *ii)* 1,000 and *iii)* 10,000 offspring from each parent, which amounts to *i)* 100,000 *ii)* 1,000,000 and *iii)* 10,000,000 total offspring, as indicated on the horizontal axes. The vertical axes show **A)** the number of distinct novel phenotypes (among a possible total of 49 phenotypes) that emerged in the offspring, and **B)** the coefficient of variation in the number of innovative offspring for different novel carbon usage phenotypes. In these analyses, parental metabolic networks contain ‖G‖=2079 reactions, the same as the *E.coli* metabolic network, and they differ in *D=100* reactions. Moreover, *n=10* reactions are swapped between parental metabolic networks during recombination.

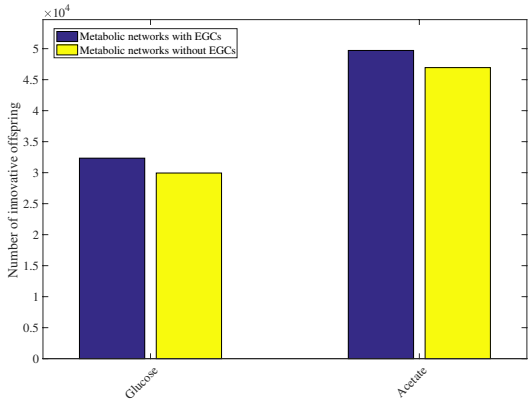
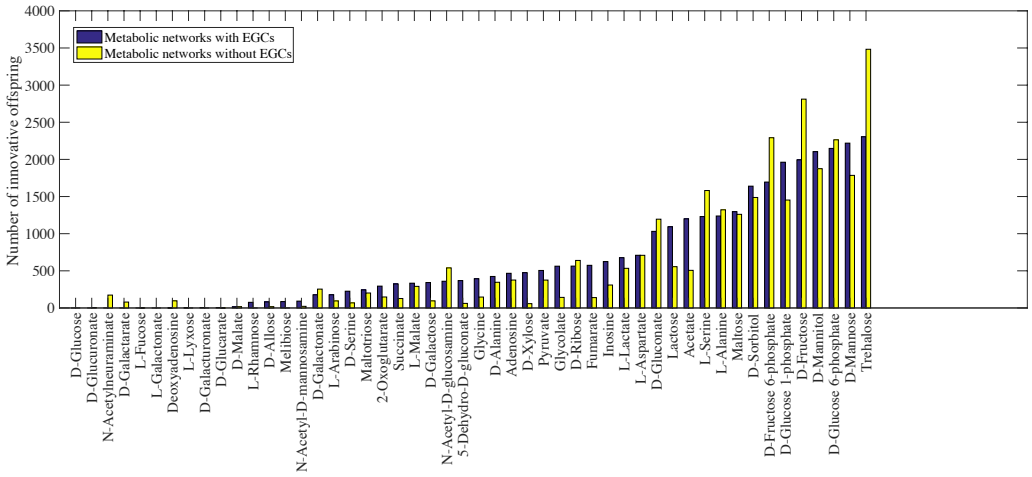


Figure S30: Erroneous energy generating cycles (EGCs) and the emergence of innovative offspring. The number of innovative offspring (per 1 million recombinants) emerging from recombination between parental metabolic networks that contain EGCs (blue) or that do not contain EGCs (yellow), and that are viable exclusively on glucose (left) and acetate (right). In these analyses, parental metabolic networks contain ‖G‖=2079 reactions, the same as the *E.coli* metabolic network, and they differ in *D=100* reactions. Moreover, *n=10* reactions are swapped between parental metabolic networks during recombination.

A)



B)

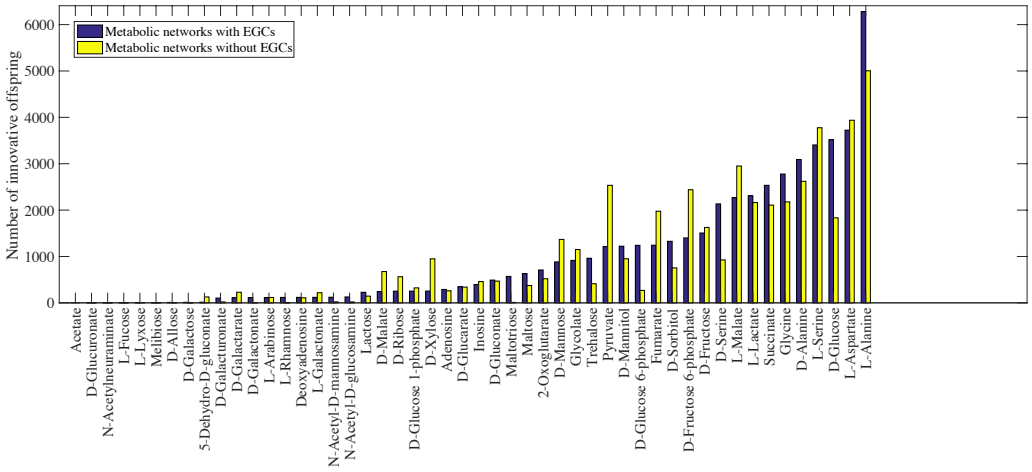


Figure S31: Erroneous energy generating cycles (EGCs) and relative constraints. Horizontal axes show the number of innovative offspring (per 1 million recombinants) emerging from recombination between parental metabolic networks viable exclusively on **A)** glucose and **B)** acetate, where parental metabolisms contain EGCs (blue) or do not contain EGCs (yellow). The ranking of the height of the blue bars and yellow bars in both panels is significantly correlated (panel A: Spearman's $\rho = 0.8913$, and $P < 10^{-18}$; panel B: Spearman's $\rho = 0.9197$, and $P < 10^{-21}$). In these analyses, parental metabolic networks contain $\|G\|=2079$ reactions, the same as the *E.coli* metabolic network, and they differ in $D=100$ reactions. Moreover, $n=10$ reactions are swapped between parental metabolic networks during recombination.

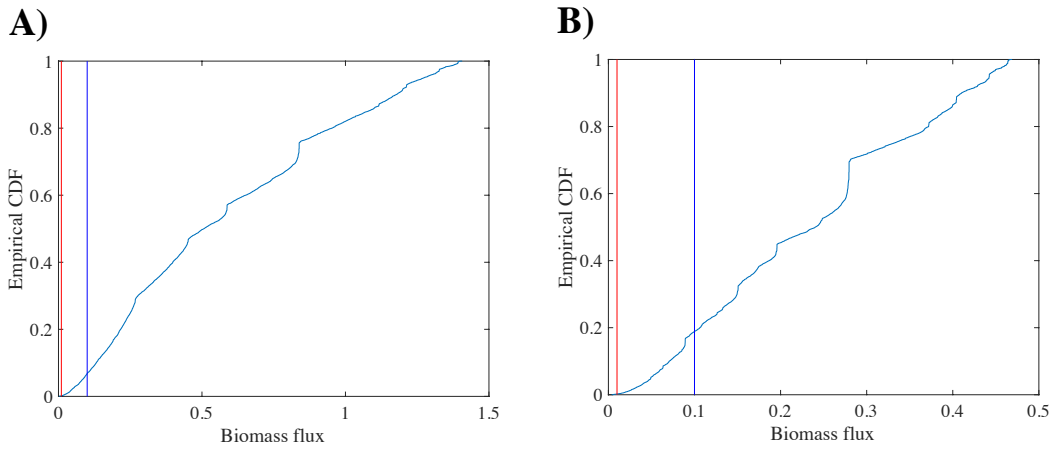


Figure S32: Biomass growth flux of most viable metabolic networks is much greater than our cut-off value for viability. The vertical axes show the empirical cumulative distribution function of the biomass flux among 10,000 MCMC-sampled metabolic networks viable exclusively on **A)** glucose, and **B)** acetate. The vertical red and blue lines show the cut-off value of 0.01 and 0.1 $1/h$. We used 0.001 $1/h$ as the cut-off value for viability.